MODELLING BUILDING CONSTRUCTION DURATIONS IN NIGERIA

By

Hassan Adaviriku AHMADU, BSc (A.B.U) 2009

(MSc/Env-Des/1515/2011-2012)

A THESIS SUBMITTED TO THE SCHOOL OF POST GRADUATE STUDIES,
AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD
OF A
MASTER DEGREE IN PROJECT MANAGEMENT.

DEPARTMENT OF QUANTITY SURVEYING,
FACULTY OF ENVIRONMENTAL DESIGN,
AHMADU BELLO UNIVERSITY, ZARIA
NIGERIA

MAY, 2014

DECLARATION

I declare that the work in this thesis entitled "Modelling building construction durations in Nigeria" has been carried out by me in the Department of Quantity Surveying, Faculty of Environmental Design, Ahmadu Bello University Zaria.

The information derived from the literature has been duly acknowledged in the text and a list of references provided. No part of this work has been presented for another degree or diploma at any Institution.

AHMADU HASSAN ADAVIRIKU	(Signature)	Date

CERTIFICATION

This thesis entitled "MODELLING BUILDING CONSTRUCTION DURATIONS IN NIGERIA" by Hassan Adaviriku AHMADU meets the regulations governing the award of the degree of M.Sc. Project Management of the Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Chairman, Supervisory Committee (Name)	(Signature)	Date
Member, Supervisory Committee (Name)	(Signature)	Date
Head of Department (Name)	(Signature)	Date
Dean, School of Postgraduate Studies (Name)	(Signature)	Date

DEDICATION

This project is dedicated to my late father, Mal. S.O.D Ahmadu. May Allah (S.W.A) bless him with Aljannah (Amin).

ACKNOWLEDGEMENT

All praise is to Allah (S.W.A) who has made it possible for me to successfully complete my pursuit for a Master's degree.

I wish to express my sincere gratitude to my supervisory team; Dr. Y.M. Ibrahim and Dr. A.D. Ibrahim, for their invaluable advice, constructive criticism and encouragement, which has led to the successful completion of this work. My sincere gratitude also goes to the weekly seminar presentation crew; Dr. A.D. Ibrahim, Dr. Y.M. Ibrahim, Mr. B.A. Kolo, Dr. K.J. Adogbo, Mr. M. Abdulrazaq and several other lecturers of the department, for the inspiration which they gave to us during the presentations and the collective effort which they made to give directions to our research work. I owe a lot of appreciation to all other lecturers and administrative staff of the department, for making sure the best of training was what I received during the course of my programme.

I must express my sincere and profound gratitude to my mother, Mrs. A. Ahmadu, for the love, prayers and financial support she offered me throughout my programme. I consider myself fortunate having her as my mother. My gratitude also goes to my brothers and sisters for the care, encouragement and good examples they set down for me to follow. They indeed serve as role models to me.

Finally, my appreciation goes to the entire members of my class, for their contributions and for making my programme worthwhile.

ABSTRACT

Following widespread acceptance by researchers that the effects of qualitative/managerial construction time-influencing factors need to be considered in project scope-based construction time predicting models, several multivariate models combining project scope and qualitative/managerial factors have been developed. However, it has been shown in literature that the applicability of these models is clearly limited to the regions/countries where they were developed. This study was therefore aimed at developing a multivariate construction time predicting model that will be applicable to the Nigeria construction industry. A self-administered questionnaire survey was used to source information on the quantitative (project scope) factors considered in the study as well as to assess the extent of influence of the qualitative factors on construction time. Principal component regression was used for the data analysis and model development, using SPSS 16.0 for windows. Following a non-normal distribution of errors and a low R^2 value obtained when multiple linear regression analysis was first conducted, the study's data set was double log transformed and then partitioned/reclassified to account for public and private sector projects. Three models were developed following the multiple linear regression analysis repeated after transforming and partitioning/reclassifying the study's data set. Two of these models (the public sector model and the private sector model) had high R^2 values and were found after testing and validation, to be suitable for predicting construction time, while one of the models (the all projects model) had a low R^2 value and was consequently found to be unsuitable for predicting construction time. The models with high R^2 values serve as a useful tool to project managers and contractors for predicting construction time, thereby facilitating effective planning.

TABLE OF CONTENTS

Page

DECLARATION i

CERTIFICATION ii

DEDICATION iii

ACKNOWLEDGEMENT iv

ABSTRACT v

TABLE OF CONTENTS vi

LIST OF FIGURES ix

LIST OF TABLES x

1.0 INTRODUCTION 1

1.1 Background to the study 1

1.2 Statement of the problem 3

1.3 Aim and Objectives 4

1.3.1 Aim of the research 4

1.3.2 Objectives 4

1.4 Justification for the Study 5

1.5 Scope and Limitations 5

1.5.1 Scope of the study 5

1.5.2 Limitations 6

2.0 LITERATURE REVIEW 7

2.1 Construction Time Estimating Methods in Practice 7

2.2 Modelling Construction Time 8

2.2.1 The Bromilow (1969) time-cost model 8

2.2.2 Attempts to validate the Bromilow (1969) time-cost model 10

2.2.3 Attempts to modify the Bromilow's time-cost model 15

2.3 Modelling construction time in Nigeria 25

2.4 Summary **29**

3.0 RESEARCH METHODS **31**

3.1 Research Approach **31**

3.2 The Quantitative Research Approach **31**

3.3 The Self-administered Questionnaire Survey **32**

3.3.3 Structure of the questionnaire 35

3.4 Statistical Methods/Approach used for Model Development **37**

3.4.1 Principal Component Analysis of the qualitative factors 39

3.4.2 Multiple linear regression analysis to determine the nature and strength of the relationship between duration and a combination of the quantitative factors and the principal components of the qualitative factors 41

3.5 Approach used for model testing and validation **45**

4.0 DATA PRESENTATION, ANALYSIS AND DISCUSSION **46**

4.1 Data Received from the Survey **46**

4.2 Data Reliability **47**

4.3 Characteristics of the Projects Surveyed **48**

4.4 Potential Predictor Variables **49**

4.5 Multivariate Model Development **50**

4.5.1 Principal component analysis of the qualitative factors 50

4.5.2 Multiple linear regression analysis to determine the nature of the relationship between construction time and construction time influencing factors 57

4.5.3 Developing the Models 75

4.5.3.1 Transforming back and interpreting coefficients 76

4.6 Model testing and validation **81**

4.6.1 Checking for the assumptions of linearity 81

4.6.2 Comparing the planner's estimates and the model's predictions 82

5.0 SUMMARY, CONCLUSION AND RECOMMENDATION **87**

5.1 Summary of findings **87**

5.2 Conclusions 90

5.3 Recommendations 90

REFERENCES 91

APPENDIX A 96

APPENDIX B 103

LIST OF FIGURES

Page

Figure 1: Rank/Position of Respondents 47

Figure 2: The Scree Plot 53

Figure 3: Error Distribution of Fitted Model 59

Figure 4: Homogeneity of Errors of Fitted Model 60

Figure 5: Identification of outliers 62

Figure 6: Error Distribution of Fitted Model 65

Figure 7: Homogeneity of Errors of Fitted Model 65

Figure 8: Error Distribution of Fitted Regression Model 69

Figure 9: Homogeneity of Errors of Fitted Regression Model 69

Figure 10: Error Distribution of Fitted Regression Model 72

Figure 11: Homogeneity of Errors of Fitted Regression Model 72

LIST OF TABLES

Page

Table 1: Regression Models Developed by Chan and Chan (2002) 24
Table 2: Models Derived by Ogunsemi and Jagboro (2006) using Double-log Linear Regression..... 26
Table 3: Models Derived by Ogunsemi and Jagboro (2006) using Piecewise Linear Regression with Break Points 26
Table 4: Factors having Stronger Effect on Construction Time 28
Table 5: The basic assumptions of linearity that justifies the use of linear regression models..... 43
Table 6: Percentage of Questionnaires Returned and Not Returned 46
Table 7: Percentage of Questionnaires Received in each Sector 46
Table 8: Information Relating to Projects Surveyed 48
Table 9: Evaluation of the Project Scope Factors Considered in the Study 49
Table 10: Analysis of the Qualitative Factors Considered in the Study 49
Table 11: The KMO and Bartlett's Test 51
Table 12: Extraction of Principal Components..... 52
Table 13: Rotated Component Matrix..... 54
Table 14: Selected Principal Corresponding Components 57
Table 15: Summary of Fitted Regression Model 58
Table 16: ANOVA Results of Regression Analysis 58
Table 17: Regression Coefficients of Fitted Model 59
Table 18: Summary of Fitted Regression Model 63
Table 19: ANOVA Results of Regression Analysis 64
Table 20: Regression Coefficients of Fitted Model 64
Table 21: Summary of Fitted Regression Model 67
Table 22: ANOVA Results of Regression Analysis 68
Table 23: Regression Coefficients of Fitted Regression Model 68
Table 24: Summary of Fitted Regression Model 70
Table 25: ANOVA Results of Regression Analysis 71
Table 26: Regression Coefficients of Fitted Regression Model 71
Table 27: Interpreting Linear Regressions with Log Transformed Variables 77

Table 28: Comparing Model Predictions to Planner's Estimates 82
Table 29: Statistical Significance of the Difference between Model's Predictions and Planner's Estimates 83
Table 30: Comparing Model Predictions to Planner's Estimates 83
Table 31: Statistical Significance of the Difference between Model's Predictions and Planner's Estimates 84
Table 32: Comparing Model Predictions to Planner's Estimates 85
Table 33: Statistical Significance of the Difference between Model's Predictions and Planner's Estimates 85

CHAPTER 1

INTRODUCTION

1.1 Background to the study

The importance of ensuring accuracy and reliability of construction time estimates at the tendering stage cannot be overemphasized. Accurate early estimates of construction time typically provides clients and contractors with a basis for evaluating the success of a project and the efficiency of the project organisation (Nkado, 1995). They also provide them with a basis for ascertaining logistical and cash flow implications for feasibility, budgeting, planning, monitoring and even litigation purposes (Skitmore and Ng, 2003). Furthermore, they serve as a criteria in determining the best combination when performing time-cost optimization (Que, 2002). It is therefore clear that construction time has become a vital tool used by clients and contractors to ensure the success of construction projects. This success will however, only be achieved when construction time is accurately predicted.

Construction time/periods are often calculated on the basis of the planner's own previous experience on similar projects (Choudhury and Phatak, 2004). However, as pointed out by Skitmore and Ng (2003), the need to reduce the problem of subjectivity associated with the planner's experience and judgement to correctly interpret project and site information and make the best possible decisions, has long resulted in the development of construction time predicting models.

The development of construction time predicting models commenced with the use of project scope factors. Project scope is a measure of project size, which can be described as construction cost, project duration, gross floor area, number of storey, building type and procurement method (Walker, 1995).

The fact that the development of construction time predicting models commenced with the use of project scope factors, can clearly be seen in the time-cost model originally developed by Bromilow (1969). It is also evident in the validation of Bromilow (1969) model by other researchers e.g. in Australia (Ireland, 1983; Yeong, 1994; Ng et al., 2001), in the UK (Kaka and Price, 1991), in Hong Kong (Chan, 1999), in Nigeria (Ojo, 2001; Ogunsemi and Jagboro, 2006), in Korea (Long and Young, 2009), as well as in Ghana (Ameyaw et al., 2012). Efforts to improve the Bromilow (1969) model are also clear indications that the development of construction time predicting models commenced with the use of project scope factors, e.g. Chan and Kumaraswamy (1995) combined cost and gross floor area (GFA) to model construction time; Love et al. (2005) on the basis that cost is a poor predictor of construction time since the final cost is not known at the initial stages of a project, modelled construction time using GFA and number of storey; Chen and Huang (2006) in a bid to finding a way to quickly estimate the duration and cost of reconstruction projects in the early planning phase where detailed information is not available, modelled construction time using cost and number of floors.

Project scope based models were found good for predicting construction time, however, as researchers (e.g.; Walker, 1995; Chan, 1999; Ng et al., 2001) rightly observed, they lacked the inclusion of several other non-project scope (qualitative) factors which influence construction time. Walker (1995) specifically noted that the principal criticism of project scope based models is that the management process is too complex to be relegated to the value of a constant conveniently described in a formula, regardless of the correctness of the statistical methodology used. It is on this basis that a recommendation for further research was generally been made by the aforementioned research works, for the identification of those other non-project scope (qualitative) factors which influence

construction time, and the full incorporation of their coefficients and weightings into a model for predicting construction time.

Several researches have been carried out to identify other non-project scope (qualitative) factors which influence construction time, some of which include the works of Ireland (1985), Walker (1995), Nkado (1995), Koushiki et al. (2005) as well as Olupolola et al. (2010). Research works have also been conducted to develop multivariate construction time predicting models combining project scope and other qualitative construction time influencing factors together (e.g., Chan and Chan (2002) developed a multivariate model for public housing projects in Hong Kong; Hoffman et al. (2007) also developed a multivariate model for facility projects funded by the US air force).

However, because there is little consensus as to which combination of the qualitative construction time influencing factors provide an accurate predictor of construction duration (Nkado, 1999) and inaccuracies may result when regression models are used to extrapolate or predict values of the dependent variable for independent variables which are outside the population for which the original model was developed (McClave et al., 2000), it is obvious that existing multivariate models are only applicable in the regions/countries where they were developed.

1.2 Statement of the problem

Project scope based construction time predicting model has been developed for Nigeria by Ogunsemi and Jagboro (2006). However, the research work of Olupolola et al. (2010) which revealed 23 key factors as having stronger effect on the time performance of building projects in Nigeria, is an indication that just like other project scope based construction time predicting models developed across the world, the time-cost model

developed by Ogunsemi and Jagboro (2006) falls short of other factors that influence construction time (in Nigeria).

Nigeria is therefore faced with the question of how these other factors can be combined with project scope factors to predict construction time, since despite the fact that multivariate models have been developed in some parts of the world they cannot be used in Nigeria because as Nkado (1995) rightly observed, there is little consensus as to which combination of these more qualitative factors provides an accurate predictor of construction duration. In addition, McClave et al. (2000) has noted that inaccuracies may result when regression models are used to extrapolate or predict values of the dependent variable for independent variables which are outside the population for which the original model was developed.1.3 Aim and Objectives

1.3.1 Aim of the research

To develop a multivariate construction time predicting model for Nigeria.

1.3.2 Objectives

The objectives through which the above stated aim was achieved were as follows:

i. To evaluate the project scope factors influencing construction time of building projects.
ii. To assess the qualitative factors influencing durations of building projects.
iii. To develop the model.
iv. To test the model.

1.4 Justification for the Study

Modelling building construction durations in Nigeria will provide Nigeria with a construction time predicting model which incorporates both project scope factors and managerial/qualitative construction time influencing factors together. This will serve as a means to ensure more accurate and realistic predictions of construction time as all construction time influencing factors will be taken into consideration any time the model is used to make predictions.

Consultancy firms and contractor organizations will therefore have the opportunity to make more realistic schedules and benchmark measures of construction period, thereby facilitating effective planning.

1.5 Scope and Limitations

1.5.1 Scope of the study

Building construction projects of a contract sum not less than ten million naira, completed via the traditional procurement route, between the years 2007-2012 were considered for this research. The range of years (2007-2012) and the choice of projects having a contract sum not less than ten million naira, was to ensure that the most recent trend within the construction industry is captured. As it was not possible to cover the whole Nigeria in the study, only Abuja (FCT), Kaduna and Kano states of Nigeria were covered by the study. These states were selected because they were considered to be among some of Nigeria's most developed states, with a relatively high concentration of construction firms in them.

1.5.2 Limitations

i. The usefulness of the models developed in this study is limited to the accuracy of the data provided by respondents in the self-administered questionnaire survey.

ii. As with any prediction model, prediction is only valid within the range of characteristics of the selected sample data. For this reason, the use of the models developed in this study, is limited to projects that meet the characteristics of the sample used for the model development i.e. public and private sector building projects of a contract sum not less than 10 million naira, completed via the traditional procurement route.

iii. Cost, GFA and number of floors were the only quantitative/project scope factors considered in the study. The developed models therefore do not account for any variation in duration which may arise from other project scope factors such as location, procurement route and type of contract.

CHAPTER 2

LITERATURE REVIEW

2.1 Construction Time Estimating Methods in Practice

According to Telford (1994) there are two common methods of estimating project completion time in practice. They include; (1) time constraints of the client e.g. occupancy need or, (2) through a detailed analysis of work to be done and resources available, using estimates of the time requirements for each specific activity. Nkado (1995) corroborated the above assertion by Telford (1994) as he noted that construction time can be deduced from the client's brief or derived by the construction planner from available project information.

Two major shortcomings have however been identified with these methods. The first method can lead to unrealistic construction time estimates caused by external factors, mostly in the form of a fixed date of occupancy (Hoffman, 2007). The second method on the other hand, could be very tedious and often impractical in view of the time limitations imposed on contractors at the tendering stage (Ng et al., 2001).

Both shortcomings point to the fact that; 1) these methods are highly dependent on opinions or feelings, rather than on facts or evidence and 2) they cannot be used when construction time estimates need to be made as quickly as possible. Attempts to develop a means by which construction durations can be accurately predicted in a quick, quantitative and objective manner has however, long become a problem of concern and interest to both researchers and practitioners in the construction industry as can be seen in the several efforts been made to model construction time.

2.2 Modelling Construction Time

Efforts to model construction time commenced largely with the use of project scope factors. This can clearly be seen in; the time-cost model originally developed by Bromilow (1969); the validation of the Bromilow (1969) model by other researchers as well as in; attempts to modify the Bromilow (1969)'s time-cost model.

2.2.1 The Bromilow (1969) time-cost model

The first model of construction time was developed by Bromilow (1969), following his pioneering investigation of the time performance of 309 building projects completed in Australia between the year 1964 and 1967. The model, often referred to as the Bromilow's time-cost (BTC) model, enables construction time to be predicted in the form of the formula:

$$T = KC^B,$$

Where T is the duration of the construction period from date of site possession to practical completion, in working days; C is the final cost of the building in millions of Australian dollars, adjusted to constant labour and material prices; K is a constant describing the general level of time performance for a \$1 million project and; B is a constant describing how the time performance is affected by project size, as measured by cost.

Bromilow (1969) considered cost to be the best indicator of project duration on the basis that; it provides not only a measure of the physical size of the project, but also reflected the complexity and quality of the work completed. Bromilow (1969) was able to establish the above relationship between construction duration and cost, through the use

of regression analysis in which he obtained the values of K and B to be 313 and 0.3 respectively. The resulting model was non-linear and can be expressed as:

$$T = 313C^{0.3}$$

In order to make the model linear and thus, obtain a more significant time-cost relationship, Bromilow (1969) transformed his data set using the logarithmic transformation. The non-linear model in the form of a power equation thus, became linear after the logarithmic transformation was applied to the data set. The resulting linear model took the form:

$$\ln(T) = \ln(KC^B) = \ln(K) + B\ln(C)$$

The Bromilow's (1969) time-cost model provided a quick and objective means of predicting construction duration. It however had a limited applicability due to the fact that each construction industry as well as project type has its own unique characteristics and therefore different parameter estimates of K and B in the Bromilow's time-cost relationship may exist for different project types as well as for different construction industries. This limitation basically provided the need for the research work conducted by Bromilow et al. (1980) which was essentially a revised version of the Bromilow (1969) time-cost model.

In Bromilow et al. (1980), the time-cost data for a total of 419 building projects completed in Australia between 1970 and 1976 were investigated. However, unlike in Bromilow (1969) where the sample data was considered wholly (despite containing significant variability in project type), in Bromilow et al. (1980) the sample data was partitioned in order to account for differences in project characteristics. It was partitioned into government projects and private sector projects completed prior to 1974 and

government projects and private projects completed after 1974. This resulted into four different models which all confirmed the initial Bromilow (1969) time-cost relationship and also, illustrated that significantly different parameter values could be used to model construction durations across a range of differing project characteristics. The models and their corresponding *K* and *B* values were as shown below:

- Pre-1974 Private: $T = 232C^{0.28}$
- Pre-1974 Government: $T = 335C^{0.28}$
- Post-1974 Private: $T = 243C^{0.37}$
- Post-1974 Government: $T = 406C^{0.34}$

2.2.2 Attempts to validate the Bromilow (1969) time-cost model

Bromilow et al. (1980)'s work showed that as anticipated, different parameter values of *K* and *B* in the Bromilow (1969) time-cost model exists for different project types and construction industries. This prompted several other researches to be carried out with the aim of validating the Bromilow (1969) time-cost model and thus, calibrate it for use across different project types and construction industries. Amongst the several researches conducted to validate the Bromilow's time-cost model was the research work conducted by Ireland (1983) to predict construction time of high rise buildings in Australia. Ireland (1983) replicated the work of Bromilow (1969) using time-cost data of 25 high rise building projects in Australia and obtained a time- cost model given by the formula:

$T = 219C^{0.47}$

Other researches undertaken to validate the Bromilow's model include;

The work of Kaka and Price (1991) which was aimed at establishing a time-cost relationship for both building and road work projects completed between 1984 and 1989 in the United Kingdom. Kaka and Price (1991) adopted the data partitioning approach

used by Bromilow et al. (1980) and they were able to establish time-cost relationships similar to that developed by Bromilow et al. (1980). Kaka and Price (1991) found that even when there is a significant variation in the estimated and actual values of construction duration, the relationship between time and cost remained highly correlated. Kaka and Price (1991) also found that both type of client and type of project influenced the time-cost relationships.

Yeong (1994) replicated the work of Bromilow (1969) to determine the time-cost relationship of building projects in both Australia and Malaysia. Making use of 67 Australian government projects, 20 Australian private projects and 51 Malaysian government projects, Yeong (1994) was able to establish the following time-cost models:

Australian private projects: $T = 161C^{0.367}$

Australian government projects: $T = 287C^{0.237}$

All Australian projects: $T = 269C^{0.215}$

Malaysian government projects: $T = 518C^{0.352}$

Kumaraswamy and Chan (1995) probed the range of factors affecting construction project duration in Hong Kong through the use of a literature review and a pilot survey in Hong Kong. They compared the data which they obtained in Hong Kong with the time-cost model developed by Bromilow et al. (1980) in Australia and found that their data set contained patterns similar to Bromilow et al. (1980)'s time-cost model although larger samples were required to validate the parameter values of their proposed model.

Chan (1999) conducted a research to determine whether the Bromilow's time-cost relationship ($T = KC^B$) holds for building projects in Hong Kong. Chan (1999) analyzed the time-cost data of 110 building projects in Hong Kong using regression analysis and confirmed that contract time and cost have a relationship of the form $T = KC^B$. The

values which he obtained for K and B in his analysis, showed that the best predictor of average construction time in Hong Kong was given by $T = 152C^{0.29}$. Chan (1999) concluded that the derived time-cost relationship expressed in the form $T = KC^{B}$, provides an alternative and objective method of estimating construction time to supplement the practice of estimation based largely on the individual project manager's experience.

Ng et al. (2001) on the basis that there is no guarantee that the parameters of the Bromilow's time-cost model or even its form will be invariant over time, conducted a research to revalidate the Bromilow's time-cost model and examine its variability over different time periods. To do this, Ng et al. (2001) collected and analyzed the time-cost data of 93 projects completed in Australia between the year 1991 and 1998. They restricted their survey to projects having a contract value more than 500,000 Australian dollars and maintained that projects below 500,000 Australian dollars were limited in scope and complexity. They also rebased all costs to the current year in which the research was carried out, using the building price index (BPI) published in the Australian construction hand book. This according to them was to account for the differences in the economic climate that prevailed during the interval of years considered in their research.

Ng et al. (2001) were able to refit the Bromilow's model with their new data set, through the use of regression analysis. They compared the newly calibrated model with the original Bromilow (1969) time-cost model and found that the K value of the newly calibrated model decreased by 35% while the B value remained almost the same. This according to them indicated an improvement in construction speed over the decades since the original BTC model was developed. Ng et al. (2001) further partitioned their data set into public sector and private sector, method of contractor selection, type of project and type of contractual arrangements. They developed separate regression models for each

category and compared the models with previously developed models. They found that public sector projects in particular exhibited greater variation over the years by up to 132%. Ng et al. (2001) concluded that in the face of changes such as this, regular revalidation of the models is necessary in order to avoid their becoming obsolete.

Choudhury and Rajan (2004) carried out a study similar to that of Ireland (1985) and Yeong (1994), as they replicated the work of Bromilow (1969) to determine the time-cost relationship of residential construction projects in Texas. Making use of a questionnaire survey, Choudhury and Rajan (2004) obtained the time-cost data of 55 building construction projects completed in Texas within a five year period. They utilized the SAS statistical program to analyse their data set and they obtained a time-cost relationship given by the formula:

$$T = 18.96C^{0.39}$$

Choudhury and Rajan (2004) concluded that their derived model indicated that the Bromilow's model holds good for the Texas construction industry at a level of significance (p-value) of <0.0001. They however noted that their study was limited to the validation of the time-cost relationship developed by Bromilow et al. (1980) and did not incorporate the implications of other factors that are likely to influence the total time required for the completion of a construction project.

While the aforementioned research works were successful in their effort to calibrate the Bromilow's time-cost model for use across different project types and construction industries, some other research works found the Bromilow's time-cost relationship not applicable to their construction industry as their efforts to calibrate the model for use in their construction industry proved unsuccessful. Among these research works include:

The study conducted by Ojo (2001) to determine the relationship between contract period and tender sum of construction projects in Nigeria. Ojo (2001) though, was able to establish a time-cost relationship given by; $T = 27C^{0.125}$, concluded that the Bromilow (1969)'s widely reported time-cost model is not suitable for Nigeria as the derived model had a low coefficient of determination and consequently, a poor predictive ability. Ogunsemi and Jagboro (2006) also attempted to validate the Bromilow (1969) time-cost model in Nigeria. However, just like Ojo (2001), Ogunsemi and Jagboro (2006) also found that the Bromilow (1969)'s time-cost model was inapplicable to the Nigeria construction industry, due to the same reason of a low coefficient of determination of the derived model.

Similar results were obtained by Long and Young (2009) in Korea and Ameyaw et al. (2012) in Ghana. Long and Young analyzed the time-cost data of 34 historical building projects in Korea. The results of their analysis revealed that a time-cost relationship is applicable to the Korean construction industry. However, further analysis according to Long and Young (2009) indicated that the original Bromilow's time-cost model was not the best fit regression form and alternative models were therefore proposed. Ameyaw et al. (2012) assessed the applicability of the Bromilow's time-cost model on building projects in Ghana. They partitioned and analyzed the time-cost data of 62 building projects completed in Ghana between the years 2000-2007, using regression analysis. The models which they derived from their analysis were as shown below:

Office: $T = 344.59C^{0.684}$

Class room: $T = 512.28C^{0.463}$

Residential: $T = 2.807C^{0.399}$

Combined: $T = 3.170C^{0.378}$

These models however, just like the one developed by Ojo (2001) in Nigeria, each had a low coefficient of determination (0.463, 0.399 and 0.378 respectively) and consequently, a poor predictive ability. These low coefficients of determination was the main reason for Ameyaw et al. (2012)'s conclusion that the original Bromilow's time-cost model is not applicable to the Ghanaian construction industry.

Based on the foregoing, it can be seen that while a vast majority of the efforts to validate the Bromilow's time-cost model were successful and actually confirmed the relationship between construction time and cost as proposed by Bromilow (1969), some other research works have shown that even though a relationship of the form proposed by Bromilow (1969) exist between construction time and cost, cost alone may not be able to account for the total variability in construction time. This however, only corroborates a major shortcoming which several researchers have identified with the BTC model, that is: it falls short of other factors which influence the completion time of construction projects.

2.2.3 Attempts to modify the Bromilow's time-cost model

The Bromilow's time-cost model has been widely criticised as lacking the inclusion of other factors which influence construction time. This was clearly stated as a limitation in some of the research works which tried to validate the BTC model, for example in Chan (1999) as well as in Choudhury and Rajan (2004). Walker (1995) specifically noted that the principal criticism of project scope based models is that the management process is too complex to be relegated to the value of a constant conveniently described in a formula, regardless of the correctness of the statistical methodology used.

This limitation of the Bromilow (1969)'s time-cost model basically provided the basis for the recommendation for further research generally been made by several researchers, for

the identification of those other factors which influence the completion time of construction projects and the full incorporation of their coefficients and weightings into a model for predicting construction time. Several researches have been carried out to identify and incorporate coefficients and weightings of these other factors which influence the completion time of construction projects. Some of these research efforts have focused basically on the identification and incorporation of additional quantitative (scope related) factors, while some others have concentrated on the identification of qualitative (management related) factors. Research efforts have also been made to incorporate coefficients and weightings of both quantitative and qualitative factors in a single model.

2.2.3.1 Quantitative (scope related) factor incorporation

Research works which have focused on the identification and incorporation of additional scope related factors includes:

The research work of Chan and Kumaraswamy (1995) which was the second phase of a research aimed at investigating the significant factors influencing the construction duration of projects in Hong Kong. Chan and Kumaraswamy (1995) had in the first phase of their research, carried out a pilot survey to explore and compare the empirical relationship between duration and cost; duration and total gross floor area; duration and total number of storeys; and any other significantly related variables in representative samples within different categories of projects completed during 1990 to 1993 in Hong Kong. They discovered patterns similar to the Bromilow's time-cost relationship in the data set derived from their pilot survey, but required a larger sample to validate the K and B parameter values.

Chan and Kumaraswamy (1995) were able to increase their data set in the second phase of their research, which they then partitioned into government projects and private sector projects. On the basis that; patterns similar to the Bromilow's time-cost relationship was discovered in the data set during the first phase of the research and also; cost itself was found to be dependent on floor area as well as on the number of storeys of a building, Chan and Kumaraswamy (1995) assumed that a relationship of the form $T = KC^B$ as proposed by Bromilow (1969) existed between construction duration and total gross floor area as well as between construction duration and number of stories.

They proposed a time-floor relationship of the form $T = L\ A^M$ and a time-storey height relationship of the form $T = F\ S^G$, where A was the gross floor area and L and M corresponds to K and B in the BTC model, while S was the number of stories and F and G also corresponded to the constants K and B in the BTC model. They evaluated both relationships and confirmed that only the time-floor relationship was significant. Making use multiple regression analysis, Chan and Kumaraswamy (1995) were able to combine time, cost and gross floor area in a single model which they represented as:

$$T = K\ C^B\ A^M$$

Chan and Kumaraswamy (1995) found the corresponding parameter values of K, B and M to be different for each category of public and private sector building projects considered in their sample. They concluded that while it was possible to adopt the models individually (that is in the form $T = KC^B$ and $T = L\ A^M$), it appears better to use the combined model incorporating both cost and gross floor area, provided the corresponding macro project parameters (cost and gross floor area) are given at the planning stage.

Love et al. (2005) conducted a similar study which just like the work of Chan and Kumaraswamy (1995), focused primarily on quantitative (project scope related) factors. They examined the influence of 5 project scope factors (namely; project type, procurement method, tender type, gross floor area and number of stories) on the time-cost performance of 161 construction projects that were completed in various Australian states.

Love et al. (2005) through the use of; multiple linear regression involving the technique of weighted least squares and; the conversion of categorical variables such as procurement method, project type and tender type using a dummy variable coding of 1s and 0s, were able to examine the influence of each project scope factor on project time-cost relationship. The results of their analysis (contrary to the findings of Chan and Kumaraswamy (1995)) revealed that both gross floor area and number of floor levels in a project influence the time performance of projects, while cost was a poor predictor of construction time since one cannot know the post construction cost at the outset of a project.

Love et al. (2005) contended that it was more sensible to predict project time based on gross floor area and number of floor levels. They suggested that project time can be predicted using the model which they expressed as follows:

$$\log(T) = 3.178 + 0.27 \log(GFA) + 0.142 \log (floor)$$

The findings of Love et al. (2005)'s work concurs with Chan and Kumaraswamy (1995)'s conclusion that a significant relationship exist between gross floor area and project duration. It however, clearly contradicts Chan and Kumaraswamy (1995)'s confirmation of cost as a good predictor of construction time and number of floors as a weak predictor of construction time.

Other research works which focused on the identification and incorporation of additional project scope factors include:

The study conducted by Chen and Huang (2006), which was aimed at developing prediction models for the cost and duration of school reconstruction projects in Taiwan. Chen and Huang analyzed 132 reconstruction projects using correlation and regression analysis. They found that floor area can be used to predict the duration of reconstruction projects in Taiwan, in the form of a cubic regression equation as well as in the form of a power regression equation similar to the time-floor area relationship ($T = LA^M$), developed by Kumaraswamy and Chan (1995).

They also found that cost can be used to predict the duration of school reconstruction projects in Taiwan, in the form of the time-cost model developed by Bromilow (1969). The models developed by Chen and Huang (2006) were represented as shown below:

Central agency: $T = 131 + 0.018A – 1.085E – 6A^2 + 3.675E – 11A^3$

Local government: $T = 42.280A^{0.1966}$

Private sector: $T = 18.085C^{0.1942}$

Forsythe et al. (2007) conducted a similar study which focused on determining the extent to which GFA and number of levels are important factors in determining the construction time in Australian detached housing projects. From an analysis of 196 comparable detached housing projects, Forsythe et al. (2007) found that GFA and number of levels correlate strongly with estimated construction time, but weakly with actual construction time. Further analysis by Forsythe et al. (2007) however, indicated that dynamically changing events during construction was responsible for the weak correlation of GFA and number of levels with actual construction time.

The foregoing clearly shows that both GFA and number of levels/floors are project scope factors which just like cost, provides a good bases for predicting construction time.

2.2.3.2 Qualitative (management related) factor identification

Amongst several research works which have been conducted to identify those other qualitative (management related) factors which influence construction time, includes:

The study conducted by Ireland (1985) to investigate the influence of managerial actions on the construction time performance of high rise commercial building projects. The results of his research revealed that client experience, form of building procurement and project organizational structure are significant managerial factors which interact in a more complex model of project time performance.

Walker (1995) carried out a similar study to investigate the factors affecting construction time performance of Australian construction projects, making use of a structured survey which involved a pilot study and 33 case study investigations. He found that the principal factors affecting construction time performance were management and client related. These factors according to Walker (1995) included: construction management effectiveness, the sophistication of the client and the client's representative in terms of creating and maintaining positive project team relationships with the construction management and design team, design team effectiveness in communicating with construction management and client's representative teams and lastly, a small number of factors describing project scope and complexity.

Nkado (1995) also conducted a similar study which was aimed at prioritizing construction time-influencing factors from the view point of those involved with construction activities on site (the contractors). He made use of a questionnaire to sample the opinion of experienced construction planners on factors which are considered in

estimating construction time. He further requested the respondents to rank the factors in order of importance, using a scale of 1 to 3. The results of his analysis showed that those factors high on the priority list were generally readily identifiable from project information and directly quantifiable by the contractor (their impact on construction time can be assessed), while those factors low on the priority list were those whose effect on construction time are not readily assessed explicitly (their influence on construction time is not within the direct control of the main contractor).

According to Nkado (1995) the factors which were high on the priority list included: client's specified sequence of completion, contractor's programming of the construction work, form of construction, client's and designer's priority on construction time, complexity of project, project location, buildability of design, availability of construction management team and completeness and timeliness of project information. On the other hand, the factors which were low on the priority list included; the form of contract and its suitability to the project, the work of statutory undertakers, previous working relationship with and provision for ease of communication with the design team, the weather, whether the client is a public or a private organization, effect of building and fire regulations as well as variations to design.

Other research works which focused on the qualitative factors that influence construction time includes:

The study conducted by Chan and Kumaraswamy (1997) which revealed; poor site management and supervision; design complexity and mistakes and delays in design documentation as well as; skills gaps, motivation gaps and productivity gaps, as the top ranked factors responsible for time overruns in Hong Kong construction projects.

In a similar study on time-delays and cost-increases associated with the construction of private residential projects in the state of Kuwait, Koushki et al. (2005) from a person-interview survey of 450 randomly selected private residential owners and developers, found that the three main causes of time-delays included; changing orders, financial constraints and owners' lack of experience in the construction business.

From the foregoing research works it can be seen that as Nkado (1995) rightly observed, researchers view the qualitative factors which influence construction time from diverse perspectives and there is no consensus on the identification of factors which influence construction times of buildings.

2.2.3.3 Quantitative and Qualitative factor incorporation (multivariate models)

While researches have been conducted to; identify and incorporate additional quantitative factors as well as; identify those other qualitative factors which influence construction time as discussed lately, research works have also been carried out to incorporate both the quantitative factors and the other more qualitative factors together in a single model for predicting construction duration. Amongst these research efforts includes:

The study conducted by Chan and Chan (2002) which was aimed at formulating benchmark measures of industry norms for overall construction period of public housing projects in Hong Kong. In order to achieve this aim, Chan and Chan (2002) modelled the durations of primary work packages of the building construction process of standard harmony type housing blocks in Hong Kong. They made use of semi-structured interviews with senior construction professionals as well as a series of case studies on projects still under construction, to reach a consensus agreement on the categorization of the common primary work packages and the work sequencing of these packages for the construction of a new public housing project. They later used a questionnaire survey to

obtain details of 84 qualitative and quantitative time influencing factors with respect to the identified work packages. Information relating to 56 standard harmony type domestic blocks was obtained from their questionnaire survey, which they analysed using multiple linear regression analysis.

Chan and Chan (2002) derived a total of 11 regression model equations to model the duration of the primary work packages of standard Harmony type housing blocks in Hong Kong. Speed of decision making involving all project teams and communications between contracting parties, were the qualitative factors found to considerably affect duration in some of the regression model equations generated, while cost, height of building, the presence/absence of precast facade and type of foundation were some of the quantitative factors found to influence duration in the developed models. Chan and Chan (2002) verified their models using additional but similar projects to test any discrepancies between the predicted values and actual estimates proposed by contractors. The results of their verification showed that the predictions of the models fell within a reasonable range of the planner's estimates. The regression equation models generated were as shown in Table 1.

Table 1: Regression Models Developed by Chan and Chan (2002)

1	Loge PILING = 1.8922 + 0.2564 log ESTCOST1 + TYPEBLK (± 0.6813 for Annex; ± 0.2792 for H1; ± 0.4117 for H2; 0 for H3) + TYPEFOUN (± 0.1041 for H-pile; ± 0.2373 for caisson; 0.0152 for bored; 0 for precast) ± 0.0850 DEPFOUND + INFORM1 (± 0.3075 for fast; ± 0.2835 for normal; 0 for slow)	6	LAG2 = ± 1.0461 + 0.7641 PILING + TYPEFOUN (0.1387 for H-pile; 0.6048 for caisson; ± 0.7410 for bored; 0 for precast) + 376.908 RATIOSUB + 7.7E-05 VOLEXCAV
2	CAP/RAFT = 4.1658 + TYPEFOUN (± 1.3520 for H-pile; ± 1.2890 for caisson; ± 2.1450 for bored; ± 1.7258 for precast; 0 for raft) + 0.0548 NOSTOREY + INFCOMM2 (0 for good; ± 0.5967 for average) + SPEED1 (± 0.0292 for high; ± 0.6275 for average; 0 for low)	7	LAG3 = 2.5559 + 0.6442 CAP/RAFT + TYPEFOUN (± 0.6491 for H-pile; ± 0.6568 for caisson; ± 0.9606 for bored; ± 1.0519 for precast; 0 for raft) + 0.0503 SUPERSTR ± 0.0013 GFA/NOSTOREY
3	SUPERSTR = 4.0511 + 0.1316 HEIGHT + FA□ ADE (0 for with fa‡ades; 1.6216 for without fa‡ades) + LABOUR (± 0.8319 for readily available; 0 for somewhat dif® cult)	8	LAG4 = ± 6.7323 + 18.9072 SUP/SER ± 0.4576 SUPERSTR + INFORM2 (± 3.2718 for fast; ± 0.1525 for normal; 0 for slow) + 0.0001 AREACLAD
4	SERVICES = ± 0.8771 + QUALITY (0 for very high; ± 0.7640 for high; ± 6.1018 for average) + 0.4860 SUPERSTR + SITEACCE (± 4.5615 for excellent; 0 for very good; 0.4427 for good; 0.9925 for dif® cult; 3.2097 for extremely dif® cult) + INFORM2 (± 3.4961 for fast; ± 1.9177 for normal; 0 for slow) + 32.0980 CLAD/GFA	9	LAG5 = 14.5059 ± 10.4065 RATIOFIN ± 0.3338 LAG4 + 0.1623 SERVICES ± 0.4295 CAP/ RAFT ± 0.7330 LAG2
5	FINISHES = ± 43.5080 + WORKTYPE (2.4568 for new; 0 for refurbish) + 0.5495 SUPERSTR + 0.0004 AREACLAD + INFCOMM1 (± 5.8644 for good; ± 4.9034 for average; 0 for low) + 17.5104 STOREYHGT	10	Loge EST_TIME = 2.6031 + 0.0834 log ESTCOST + FA□ ADE (0 for with fa‡ades; 0.0497 for without fa‡ades) + 0.0024 HEIGHT + NATSITE (0.2352 for level; 0.2221 for built-up; 0 for sloping) + TYPESCH (± 0.0453 for purchase; 0 for rental)
		11	Loge ACT_TIME = 3.0264 + 0.1236 log ACT± COST + TYPESCH (± 0.0544 for purchase; 0 for rental) + FA□ ADE (0 for with fa‡ades; 0.0666 for without fa‡ades) + 1.3E ± 06 VOLTOTAL ± 0.0003 GFA/NOSTOREY

Source: Chan and Chan (2002)

Hoffman et al. (2007) also conducted a study which focused on the incorporation of qualitative and quantitative factors to predict construction duration. Their study which was aimed at gaining an insight into the significant factors impacting construction duration, involved an investigation into facility projects funded by the US air force. They obtained data relating to 616 facility projects from an existing data base and considered factors which fell within project scope and project environment categories.

Hoffman et al. (2007) analysed their data set using double log regression adopted by Bromilow (1969) as well as multiple linear regression analysis. The model which they derived via the use of the double log regression analysis had a low coefficient of determination ($R^2 = 0.337$) and was given by: $T = 26.8C^{0.202}$. Similarly, the model which they derived via the use of multiple linear regression analysis, also had a low coefficient

of determination (R^2 = 0.374) and was expressed as: $Y = 3.44 + 0.198X_1 - 0.059X_2 - 0.070X_3 - 0.222X_4 - 0.193X_5 - 0.0146X_6$,

Where Y = project duration (in days), X_1 = project cost, while the remaining X_i values represent dummy variables for the project environment factors considered in their study.

Hoffman et al. (2007) considered the coefficient of determination of the multiple linear regression model to be reasonable, attributing the low R^2 value to the nature of the projects which they considered as well as the fact that the projects varied widely in location, design and administrative procedures. Just like Chan and Chan (2002), Hoffman et al. (2007) validated the multiple linear regression model by comparing the predicted values with the planning estimates specified prior to construction. They concluded that the multiple linear regression model provided the most acceptable prediction.

2.3 Modelling construction time in Nigeria

Modelling construction time in Nigeria commenced with the study conducted by Ojo (2001) to determine the relationship between contract period and tender sum of construction projects in Nigeria. In order to determine this relationship, Ojo (2001) attempted to validate the Bromilow (1969) time-cost model in Nigeria. Ojo (2001) though, was able to establish a time-cost relationship given by; $T = 27C^{0.125}$, concluded that the Bromilow (1969)'s widely reported time-cost model was not suitable for Nigeria as the derived model had a low coefficient of determination and consequently, a poor predictive ability.

Ogunsemi and Jagboro (2006) conducted a similar study to explore a time-cost relationship that will be suitable for predicting project duration in Nigeria. They obtained the time and cost data of 87 building projects completed within a ten year period (from 1991 to 2002), from consulting quantity surveyors. They first of all, employed the

double-log linear regression adopted by Bromilow (1969), to analyze their data set and later, repeated the analysis using the piecewise linear regression with break point.

When the double-log linear regression was used to analyze the data set, Ogunsemi and Jagboro (2006) found that even though some assessment criteria such as the F-ratio and root mean square favoured the derived models, the R^2 value (coefficient of determination) of the models were very low, corroborating the earlier research carried out by Ojo (2001). However, when the analysis was later repeated, using the piecewise linear regression with break point, Ogunsemi and Jagboro (2006) found the resulting models to be suitable for predicting construction duration, as the coefficient of determination of the models were all very high. The models derived from each method were as shown below.

Table 2: Models Derived by Ogunsemi and Jagboro (2006) using Double-log Linear Regression

	Model	R	R^2
All projects	$T = 63C^{0.262}$	0.453	0.205
Private projects	$T = 55C^{0.312}$	0.443	0.196
Public Projects	$T = 69C^{0.255}$	0.567	0.322

Source: Ogunsemi and Jagboro (2006)

Table 3: Models Derived by Ogunsemi and Jagboro (2006) using Piecewise Linear Regression with Break Points

	Model	R	R^2
All projects	$T = 118.563-0.401C(C \leq 408)$ or $603.427 + 0.610C$ $(C>408)$	0.875	0.7656
Private projects	$T = 168.895 + 0.491C(C \leq 557)$ or $709.66 + 0.884C$ $(C>557)$	0.881	0.7762
Public Projects	$T = 98.010 + 0.357C(C \leq 353)$ or $567.967 + 0.283C$ $(C> 353)$	0.9114	0.8306

Source: Ogunsemi and Jagboro (2006)

Following Ogunsemi and Jagboro (2006)'s research work, was a study conducted by Olupolola et al. (2010), on the factors affecting the time performance of building projects in Nigeria. Olupolola et al. (2010) carried out an extensive literature review and identified 74 variables/factors which were likely to affect time performance of building projects. They later utilized a questionnaire survey to investigate construction professionals (architects, builders, engineers, project managers and Quantity surveyors) on the effects which the identified factors had on construction time performance of building projects in Nigeria.

They made use of descriptive statistics to analyse the data which they obtained from the questionnaire survey and ranked the factors/variables in relation to the extent to which they were perceived by the professionals to affect construction time. Based on the results of the analysis carried out, Olupolola et al. (2010) concluded that 23 key factors out of the 74 factors identified from the review of literature, had stronger effect on the time performance of building projects in Nigeria. Amongst these 23 key factors, delay in subcontractors work and ineffective planning and scheduling of project by the contractor were the first most important factors affecting construction time, while improper construction methods implemented by the contractor was the second most important factor affecting construction time. Table 4 shows a complete list of these factors as outlined by Olupolola et al. (2010).

Table 4: Factors having Stronger Effect on Construction Time

Factors	Rank
Delay in sub-contractors work	1
Ineffective planning and scheduling of project by contractor	1
Improper construction methods implemented by contractor	3
Conflicts in sub-contractors schedule in execution of project	4
Rework due to errors during construction	5
Late procurement of materials	6
Lateness in revising and approving design document by owner	6
Change order by owner during construction	8
Delay in progress payment by owner	9
Poor communication and coordination by contractor with other parties	10
Difficulties in financing project by contractor	10
Delay in material delivery	12
Delay in obtaining permits from regulatory authorities	12
Poor site management and supervision by contractors	12
Poor communication and coordination by owner with other parties	12
Delay in approving shop drawing and sample materials	12
Inadequate definition of substantial completion	12
Delay in approving major changes in the scope of work by consultant	18
Frequent change of sub-contractors because of their inefficient work	18
Poor communication/coordination between consultant & other parties	18
Rain effect on construction activities	18
Conflicts between contractor and other parties (consultant and owner)	24
Delay in delivering the site to the contractor by the owner	27
Slowness in decision making process by the owner	29
Shortage of equipment	30
Shortage of labourers	31
Unclear and inadequate drawings	31
Equipment breakdowns	33
Ineffective delay penalties	33
Changes in government regulations/laws	33
Inadequate experience of consultant	39
Low level of equipment operators skill	39
Insufficient data collection and survey before design	39
Delays in producing design documents	39
Poor qualification of the contractors technical staff	44
Mistakes and discrepancies in design documents	44
Inflexibility (rigidity) of consultant	44
Complexity of project design	48
Lateness in selection of finishing materials	48
Low productivity of labourers	50
Suspension of work by owner	51
Lack of high technology mechanical equipment	53
Delay in performing inspection & testing by consultant	53
Delay in providing services from utilities (such as water, electricity, e.t.c)	53
Inadequate design team experience	53

Source: Olupolola et al. (2010)

2.4 Summary

Modelling of construction time commenced with the use of project scope factors as evident in; the time-cost model originally developed by Bromilow (1969); the validation of the Bromilow (1969) model by other researchers as well as in; attempts to modify the Bromilow (1969)'s time-cost model. Project scope based construction time predicting models though held good for predicting construction time, generally lacked the inclusion of several other qualitative/managerial factors which influenced construction time. As a result of this shortcoming, several researchers made a recommendation for further studies, for the identification of those other managerial/qualitative factors and the full incorporation of their coefficient and weightings into a model for predicting construction duration.

Several research works have been carried out to identify these other more qualitative time influencing factors, researches have also been carried out to develop multivariate models combining both quantitative (project scope related) factors and qualitative (management related) factors, with these models also found to be good for making predictions of construction time.

In Nigeria, a time-cost model was successfully calibrated for the Nigeria construction industry by Ogunsemi and Jagboro (2006). However, a later research work conducted by Olupolola et al. (2010) revealed 23 key factors as having a stronger effect on the time performance of building projects in Nigeria, indicating clearly that just like other project scope based models developed across the world, the time-cost model developed by Ogunsemi and Jagboro falls short of several other factors which influence construction time in Nigeria. Nigeria is therefore faced with the question of how these other factors can be combined with the project scope factors to predict construction time, since despite

the fact that multivariate models have been developed in some parts of the world they cannot be used in Nigeria because as Nkado (1995) rightly observed, there is little consensus as to which combination of these more qualitative factors provides an accurate predictor of construction duration. In addition, McClave et al. (2000) has noted that inaccuracies may result when regression models are used to extrapolate or predict values of the dependent variable for independent variables which are outside the population for which the original model was developed.

CHAPTER 3

RESEARCH METHODS

3.1 Research Approach

In order to achieve the research objectives set out in chapter one and thus, realise the aim of the research, a quantitative research approach involving the use of a self-administered questionnaire survey was adopted for the research. As defined by Hughes (2006), quantitative research is concerned with the collection and analysis of data in numeric form. The self-administered questionnaire survey yielded the required numeric data which was then used for the model development and testing.

3.2 The Quantitative Research Approach

Quantitative research approaches seek to gather factual data and to study relationships between facts and how such facts and relationships accord with theories and findings of any research executed previously (in literature) (Fellows and Liu, 1999).

It is usually driven by the researcher's concerns and deals with measurable and quantifiable aspects of phenomena under study. It focuses on the questions; to what extent? By how much? What relationship exists between factors? What causes particular processes or situations? It is concerned with phenomena amenable to measurement of quantity and amount (Eboh, 2009). In quantitative research approaches, scientific techniques are used to obtain measurements (quantified data) e.g. social surveys such as; self-administered questionnaires, interview surveys, telephone surveys, computer assisted surveys, as well as web based survey.

As stated earlier, the quantitative research approach was adopted in this study in order to achieve the research objectives and thus realise the overall research aim. In line with

previous researches (outlined in chapter 2), a self-administered questionnaire survey was the technique used to obtain measurements (quantified data).

3.3 The Self-administered Questionnaire Survey

A self-administered questionnaire survey is a survey research method which utilizes a standardized set or list of questions given to individuals or groups, the results of which can be consistently compared and contrasted. It offers researchers the opportunity to reach a large number of potential respondents in a variety of locations, especially by using mail-response or online questionnaires (Russell, 2006).

Among the types of information that can be collected by means of a questionnaire are facts, activities, level of knowledge, opinions, expectations and aspirations, membership of various groups, and attitudes and perceptions (Siniscalco and Auriat, 1998).

For the purpose of this research, the self-administered questionnaire survey was used basically to accomplish the first two objectives of the research that is; to evaluate the project scope factors influencing construction time of building projects and; to assess the qualitative factors influencing durations of building projects. Information relating to the study's unit of analysis i.e.: building construction projects, was therefore the type of information which the self-administered questionnaire survey was used to collect in this study.

3.3.1 Survey respondents and study area

In view of the nature of the information that was required to meet the objectives which the self-administered questionnaire survey was used to accomplish, building construction firms were chosen as the target respondents for the survey, while Abuja (FCT), Kaduna and Kano states of Nigeria were chosen as the study area.

Construction firms were considered better placed to provide information relating to both objectives on the basis that; they have a better and more accurate knowledge about how any given building project fared, since they were directly responsible for the construction of the project. On the other hand, Abuja (FCT), Kaduna and Kano states of Nigeria were chosen as the study areas for the survey on the basis that these states are amongst the major cities in Nigeria where most construction firms have their headquarters situated. Distributing questionnaires to the construction firms in these states will therefore approximate to obtaining details of projects completed in most parts of Nigeria.

3.3.2 Survey sampling

The objective of sampling is to provide a practical means of enabling the data collection and processing components of research to be carried out while ensuring that the sample provides a good representation of the population (Fellows and Liu, 1999). Several formulas have been developed for determining the sample size needed to be representative of a given population. Some of which includes;

- Krejcie and Morgan (1970) formula;

$$S = \frac{X^2 NP(1-P)}{d^2(N-1) + X^2 P(1-P)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where;

S = required sample size

X^2 = table of value of chi-square for 1 degree of freedom at the desired confidence level (3.841)

N = the population size

P = the population proportion (assumed to be 0.50)

d = degree of accuracy expressed as a proportion (0.05)

- Fellows and Liu (1997) formula:

$$n = \frac{\gtrsim^2 \times s^2}{(\bar{\mu} - \bar{x})^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

Where:

n = sample size

$\gtrsim$ = confidence level required

s = sample standard deviation

$\bar{\mu}$ = upper and lower confidence limits of the estimate of the population mean

$\bar{x}$ = sample mean

For the purpose of this research though, determining the sample size using any of the aforementioned formulas was not possible, due to the unknown population of building construction projects in Nigeria. However, taking a cue from Mendenhall (1993) and Leedy (1997)'s assertion that; 'a required minimal sample size of 30 was regarded as sufficiently large to provide an effective normal approximation as a general rule of thumb, regardless of the shape of the population frequency distribution', a minimum of 90 samples was set as the sample size for this research (30 samples for public and private sector residential projects, 30 samples for public and private sector educational projects and 30 samples for public and private sector commercial projects).

Also, based on Salkind (1997)'s recommendation that "if you are mailing out surveys or questionnaires, count on increasing your sample by 40%-50% to account for lost mail and uncooperative subjects", a total of 135 questionnaires were issued out to respondents, instead of the minimum of 90 set as the sample size. As it was envisaged that not all construction firms will be able to provide the kind of information requested for in the survey questionnaire, a purposive sampling survey was used to select construction firms who were issued with questionnaires through mails and administration on site.

A survey package comprising of a covering letter, survey instructions and a few separate sets of self-administered questionnaires and stamped self-addressed return envelopes was mailed to selected building construction firms whose full addresses were known. Other selected building construction firms whose addresses were not known, were approached on construction sites and served with questionnaires after they had been briefed about what the research aim was.

Taking cognizance of the point raised by Kaka and Price (1993) that; poor classification of projects has been indicated to be one of the main reasons for the unsuccessful results of some previous models, construction firms were in the survey questionnaires, restricted to building construction projects of a contract sum not less than ten million naira, completed by their construction firm within the period of 2007 to 2012, using the traditional procurement route. Building construction projects of a contract sum less than ten million naira were considered to be limited in scope and complexity, while the range of years chosen was to ensure that the most recent trend in the time performance of building projects was captured in the model to be developed.

Out of a total of 135 questionnaires administered, only 96 were completed and returned, while the remaining 114 were not returned. However, since the number of returned questionnaires exceeded the required minimum of 90 samples set as the sample size of the study, the 96 returned questionnaires were considered sufficient enough to provide a good representation of the population.

3.3.3 Structure of the questionnaire

The questionnaire was structured to source information on various quantitative and qualitative factors which have been found by previous researches to influence building construction completion time. The quantitative factors considered were basic project

scope related factors, namely; cost, gross floor area (GFA) and number of floors/storey, while the qualitative factors considered where the 23 key construction time influencing factors identified by the research work of Olupolola et al. (2010).

The three project scope related factors were considered on the basis that previous researches have not only found them to be factors which influence construction time, but have also shown them to have good predictive abilities. The 23 qualitative factors on the other hand were considered on the basis that out of several other qualitative factors which have been identified in literature, these 23 were the ones which were found by the research conducted by Olupolola et al. (2010) to particularly influence construction time of building projects in Nigeria.

Two sections (A & B) were created in the questionnaire. The first section comprised closed questions requesting for the position/rank held by the respondent as well as the respondent's number of years of experience in practice. These questions were meant to give an insight into how reliable the information supplied in the returned questionnaires were, as it was expected that professionals/personnel at the top or middle level management of any given firm will have a more comprehensive knowledge about how projects executed by the firm proceeded. The second section of the questionnaire comprised of a table partitioned into two parts (part A & B). Respondents were requested to use any building project of their choice which had a contract sum not less than ten million naira and which was completed by their construction firm within the period of 2007 to 2012, using the traditional procurement route, to complete the table. In the first part of the table (part A), respondents were required to give basic details of the project which they have decided to use, to complete the table. This basic project details included; the initial and final contract sum of the project, the initially estimated duration in which the project was to be completed and the actual/final duration taken to complete the

project, gross floor area and number of floors contained in the project, the building type and year in which the project was executed, the sector to which the project belonged to.

In the second part of the table (part B), a list of the qualitative construction time influencing factors identified by the research work of Olupolola et al. (2010) was provided together with a likert scale which ranged from 0 to 5. Respondents were requested to give their assessment (using the scale provided) of the level to which each of these qualitative factors influenced the initially estimated duration of the project, to give rise to the actual/final duration taken to complete the project. The lower limit of the scale (0) represented no influence while the upper limit (5) represented very high influence.

Part A of the table was thus, the means through which the information required to meet the first objective of the research was captured by the questionnaire, while part B of the table was the means through which the information required to meet the second objective of the research was also, captured by the questionnaire.

3.4 Statistical Methods/Approach used for Model Development

All models contain parameters (variables) which must be identified and quantified for use in the model, together with their inter-relationships (Fellows and Liu, 1999).

Following the accomplishment of the first two research objectives via the self-administered questionnaire survey just described, descriptive statistics was used to gain an overview of the quantitative data derived from these objectives, while principal component regression was adopted for the data analysis and model development, making use of SPSS 16.0 for windows as the statistical tool for carrying out the analysis. Only 90% of the data obtained from the survey was used for the model development, the remaining 10% was set aside and used for testing the derived model.

Principal component regression is a method that combines principal component analysis and multiple linear regression analysis. It establishes a relationship between the output variable (y) and the selected principal components of the input variables (x_i) (Ul-Saufie et al., 2011). Just like several other research works which have formulated prediction models using principal component regression (e.g. the work of Trost and Oberlender (2003) as well as Dascalu and Cozma (2008)), this study opted to use principal component regression, basically to avoid the problem of multicollinearity often associated with using multiple linear regression to formulate prediction models.

Multicollinearity as defined by Trost and Oberlender (2003) is a condition wherein one or more of the independent variables can be approximated by a linear combination of other independent variables. According to Field (2005), when questions are designed to measure the same underlying dimension (or dimensions), it should be expected that they will correlate with each other (since they are measuring the same thing). For the purpose of this study the likely occurrence of multicollinearity was anticipated amongst the qualitative factors as they were all used to address the same question in the self-administered questionnaire and so deemed to be closely related. The study thus, used principal component analysis to identify a reduced set of uncorrelated principal components which can summarize the qualitative factors without much loss of information in the process and thus, prevent the likely occurrence of multicolinearity when multiple linear regression analysis is used to determine the nature and strength of the relationship between the dependent variable (duration) and the explanatory variables (a combination of the quantitative and qualitative variables).

3.4.1 Principal Component Analysis of the qualitative factors

Principal component analysis (PCA) is a traditional multivariate statistical method commonly used to reduce the number of predictive variables and solve the multicolinearity problem. It looks for a few linear combinations of the variables that can be used to summarize the data without losing too much information in the process (Maitra and Yan, 2008).

In this study, the principal component analysis of the qualitative factors involved 3 main steps, namely; testing for the appropriateness of using PCA; extraction of factors/principal components and; selecting the principal components to be used in the regression analysis.

3.4.1.1 Testing for the appropriateness of using PCA

Prior to performing a principal component analysis, it is always important to check if distinct and reliable factors/components will be derived from the PCA. This according to Field (2005) is usually achieved via the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett's test both measures the appropriateness of using principal component analysis in terms of the correlation matrix of a given data set.

The KMO varies between 0 and 1. According to Field (2005), Kaiser (1974) recommends accepting; values greater than 0.5 as acceptable; values between 0.5 and 0.7 as mediocre; values between 0.7 and 0.8 as good; values between 0.8 and 0.9 as great and values above 0.9 as superb, while for values below 0.5, additional data needs to be collected as the use of PCA will be inappropriate. Similarly, the Bartlett's test of sphericity tests the null hypothesis that the original correlation matrix is an identity matrix (a matrix in which all the correlation coefficients are zero, except the diagonal

which will be one all through). The significance level of this test will therefore determine if PCA will be appropriate or not. That is the use of PCA will only be appropriate if the test is significant at a p-value < 0.05.

For the purpose of this study, the KMO was 0.76 while the Bartlett's test was highly significant ($p < 0.001$), indicating that using PCA for the qualitative factors was appropriate and also, distinct and reliable components/factors will be derived from the PCA.

3.4.1.2 Extraction of factors/principal components

Principal component analysis (PCA) extracts form a set of *p* variables, a reduced set of *m* principal components or factors that accounts for most of the variance in the *p* variables (Wuensch, 2012). The eigenvalues associated with each variable/linear component in the set of p variables as well as the sree plot, is usually used as a guide to determine the reduced set of m principal components or factors that account for most of the variance in the set of p variables.

The eigenvalue associated with each variable/linear component represents the variance explained by that particular variable/linear component. Variables/linear components with eigenvalues greater than 1, is what most authors recommend to be accepted as the reduced set of m principal components which accounts for most of the variance in the p variables. The scree plot on the other hand is a plot with eigenvalues on the ordinate and component number on the abscissa. The point of inflection at which this plot begins to tail off, signifies the point at which variables/linear components with eigenvalues greater than 1 terminates/end (Field, 2005).

For the purpose of this study, 10 out of the 45 qualitative factors (variables/linear components) had eigenvalues greater than 1, signifying that a total number of 10

principal components accounted for the total variance in the 45 variables/linear components.

3.4.1.3 Selecting principal corresponding components to be used in regression analysis

Following the extraction of the ten principal components, the final step was to examine those variables/linear components which have loaded into each principal component and then select the ten most significant variables/linear components to be used in the multiple linear regression analysis. In order to do this, the criterion for selection used in the research work of Dascalu and Cozma (2007) was adopted for this purpose. This criterion involved selecting the predictor (variable/linear component) with the biggest correlation coefficient in the rotated component matrix. In addition to this, where two predictors had very close correlation coefficient, the principal component was the predictor with smaller correlation coefficients with the other principal components from the matrix. Going by this selection criterion, the 10 principal corresponding components were selected for use in the regression analysis.

3.4.2 Multiple linear regression analysis to determine the nature and strength of the relationship between duration and a combination of the quantitative factors and the principal components of the qualitative factors

As outlined by Landau and Everitt (2004), the multiple regression model for a response variable, y, with observed values, y_1, y_2, …, y_n (where n is the sample size) and q explanatory variables, x_1, x_2, …, x_q with observed values, x_{1i}, x_{2i}, …, x_{qi} for $i = 1, …, n$, is given by:

$$y= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} +……+ \beta_q x_{qi} + \varepsilon_i$$

Where:

y = dependent or response variable (variable to be modelled),

x_1, x_2, x_n = independent variables (explanatory variables),

ε_i = residual or random error component for individual i, and

β_i = respective regression coefficients.

Four principal assumptions justify the use of linear regression models for purposes of prediction, namely:

- linearity of the relationship between dependent and independent variables
- independence of the errors (no serial correlation)
- homoscedasticity (constant variance) of the errors
- normality of the error distribution.

If any of these assumptions is violated (i.e., if there is nonlinearity, serial correlation, heteroscedasticity, and/or non-normality), then the forecasts, confidence intervals, and economic insights yielded by a regression model may be (at best) inefficient or (at worst) seriously biased or misleading (Nau, 2005).

Resulting regression models must thus, be tested for each of these assumptions and in the event where any of these assumptions is being violated, appropriate measures must be used to fix the problem. Nau (2005) outlined how the violation of each of these assumptions can be detected/tested for and how they can be fixed if found to be present. This is as summarized and shown in Table 5 below.

Table 5: The basic assumptions of linearity that justifies the use of linear regression models

Assumption to be tested for	How to detect	How to fix
Violations of linearity (nonlinearity)	Nonlinearity is usually most evident in a plot of the observed versus predicted values or a plot of residuals versus predicted values, which are a part of standard regression output. The points should be symmetrically distributed around a diagonal line in the former plot or a horizontal line in the latter plot. A "bowed" pattern/parabolic curve in any of these plots indicate nonlinearity.	Application of a nonlinear transformation to the dependent and/or independent variables, e.g a log transformation.
Violations of independence (serial correlation)	Usually indicated by: a residual autocorrelation plot in the range 0.2 to 0.4, or a Durbin-Watson statistic between 1.2 and 1.6 (for minor cases) and; an autocorrelation more negative than -0.3 or DW stat greater than 2.6 (for more significant cases)	Reconsider the transformation to the dependent and/or independent variables
Violations of homoscedasticity (heteroscedasticity)	Usually shown in a plot of residuals versus predicted value, by residuals that are getting larger (more spread out)	Heteroscedasticity is a byproduct of a significant violation of the linearity and/or independence assumptions, in which case it may also be fixed as a byproduct of fixing those problems.
Violations of normality (non-normality)	A bow-shaped pattern of deviations from the diagonal of a normal probability plot of the residuals indicates that the residuals have excessive skewness (i.e., they are not symmetrically distributed, with too many large errors in the same direction)	Nonlinear transformation of variables

Source: Nau (2005)

For the purpose of this study, based on an assumption that a multiple linear relationship of the form; $y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots\ldots + \beta_q x_{qi} + \varepsilon_i$, existed between construction duration and a combination of the quantitative and qualitative factors, a multiple linear regression analysis was performed between construction duration (as the dependent/response variable), and a combination of the three quantitative factors (cost, GFA, and number of floors) with the ten principal components of the qualitative factors, (as the independent/explanatory variables), using a forward selection procedure at a significance level of 5%. Prior to performing the multiple linear regression analysis, all

costs were rebased to prices at the last quarter of the year 2012, using the composite consumer price indices published by the Central Bank of Nigeria. This was to take care of the fact that the data collected for the study were related to projects completed at different points in time and possibly different economic conditions.

The forward selection procedure is a method that starts with a model containing none of the explanatory variables. In the first step, the procedure considers variables one by one for inclusion and selects the variable that results in the largest increase in R^2. In the second step, the procedure considers variables for inclusion in a model that only contains the variable selected in the first step. In each step, the variable with the largest increase in R^2 is selected until, according to an *F*-test, further additions are judged to not improve the model (Landau and Everitt, 2004).

In this study the forward selection procedure was iterated through five steps, a variable was selected for inclusion in the model in the first four steps, while in the fifth step, none of the remaining variables lead to an increase in R^2 in accordance with the chosen selection criteria of a 5% significance level and so the process was stopped at this fifth step. Because the results obtained from this first regression analysis indicated that; there was the presence of an outlier, the error distribution was not normal and the R^2 value was low, the regression analysis was later repeated after; the outliers had been investigated and eliminated and the study's data set transformed (in order to normalize the error distribution) and partitioned (in order to obtain a higher R^2 value).

3.5 Approach used for model testing and validation

Only when a model appears adequate in the light of the data should the fitted regression equation be interpreted (Landau and Everitt, 2004).

Three models resulted from the later regression analysis and they were each tested and validated by:

i. Checking for the assumptions of linearity made in the multiple linear regression analysis modelling process.

ii. Comparing the planner's estimates and the model's estimate for the 10% data that was set aside during the model development process.

CHAPTER 4

DATA PRESENTATION, ANALYSIS AND DISCUSSION

4.1 Data Received from the Survey

Table 6: Percentage of Questionnaires Returned and Not Returned

Questionnaires	Number	Percentage
Completed & returned	96	71
Not returned	39	29
Total	135	100

Table 7: Percentage of Questionnaires Received in each Sector

Sector	Number	Percentage
Public	59	62
Private	37	38
Total	96	100

As shown in Table 6, a total of 135 questionnaires were distributed in the survey, out of which only 96(71%) were completed and returned while the remaining 39(29%) were not returned. Also, as shown in Table 7, out of the 96 completed and returned questionnaires, 37(38%) corresponded to private sector projects, while 59(62%) were public sector projects.

Taking into account, the sample size for the study which was set at a minimum of 90 samples (30 samples for public and private sector residential projects, 30 samples for public and private sector educational projects and 30 samples for public and private sector commercial projects), the 96 questionnaires completed and returned were thus, considered to be sufficient for the data analysis.

4.2 Data Reliability

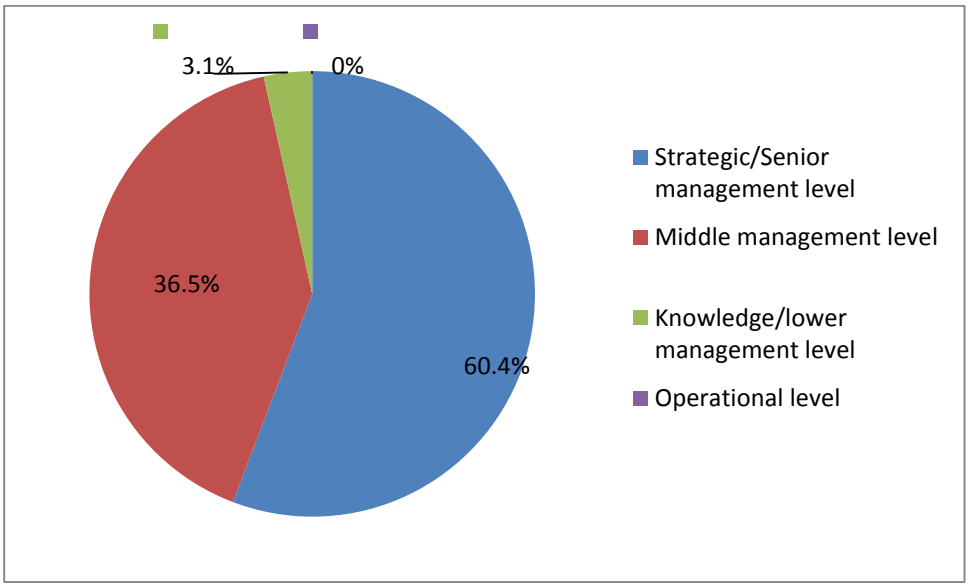


Figure 1: Rank/Position of Respondents

Data reliability is related to data source and the identification of the position held by the person who completed the questionnaire (Love et al., 2005). For the purpose of this research, it was important that professionals at the strategic/senior management level or at the middle management level of each construction firm filled the questionnaires, as they were considered to have a better knowledge of how projects executed by their firms proceeded.

From Figure 1 however, it can be seen that 60.4% of the respondents were professionals at the strategic/senior management level, 36.5% were professionals at the middle management level, and 3.1% were professionals at the knowledge/lower management level, while none of the respondents belonged to the operational level. This thus, gives a reasonable degree of reliability to the information provided by the respondents in the questionnaires.

4.3 Characteristics of the Projects Surveyed

Table 8: Information Relating to Projects Surveyed

Sector	Building type	Number	Percentage
Public	Residential	3	5.1
	Commercial	7	11.9
	Educational	45	76.2
	Health care	4	6.8
	Others	0	0
	time overrun: > 20%	59	100
	10 to 20%	0	0
	0 to 10%	0	0
Private	Residential	28	75.7
	Commercial	9	24.3
	Educational	0	0
	Health care	0	0
	Others	0	0
	time overrun: > 20%	26	70.3
	10 to 20%	5	13.5
	0 to 10%	6	16.2

Source: Field work 2013

As shown in Table 8, the building projects whose details were obtained from the survey fell under two categories; public sector and private sector projects. 76.2% of the public sector projects were educational projects, while; residential (5.1%), commercial (11.9%) and health care (6.8%), accounted for the remaining 23.8% of the total public sector projects whose details were provided. The private sector projects were mainly residential projects (75.7%), while commercial projects accounted for the remaining 24.3% of the total private sector projects obtained from the survey.

Table 8 also shows that public sector projects were more prone to time overruns of a higher degree (above/greater than 20%), compared to the private sector projects which had time overruns as low as between 0 to 10% high.

4.4 Potential Predictor Variables

For the purpose of this study, two categories of factors which influence construction time were considered for developing the proposed prediction model. These categories of factors were; the quantitative (project scope related) factors and the qualitative (management related) factors. Information obtained from the field survey with regard to these factors, were as shown in Table 9 and table 10.

Table 9: Evaluation of the Project Scope Factors Considered in the Study

Scope factor	Minimum value	Maximum value	Mean
Cost (Naira)	10,338,500	9,600,000,000	818,504,974
Gross Floor Area (m^2)	85	7,600	1,200
Number of floors (nrs.)	0	5	1

Table 10: Analysis of the Qualitative Factors Considered in the Study

Qualitative factors	Respondents' assessments						Mean
	0	1	2	3	4	5	
Delay in sub-contractors work	27	16	15	14	11	3	1.71
Ineffective planning and scheduling of project by contractor	17	11	11	22	17	8	2.41
Improper construction methods implemented by contractor	16	17	18	19	12	4	2.07
Conflicts in sub-contractors schedule in execution of project	22	20	14	14	10	6	1.86
Rework due to errors during construction	21	21	24	6	8	6	1.73
Late procurement of materials	15	8	19	32	9	3	2.24
Lateness in revising and approving design document by owner	1	21	28	14	14	8	2.5
Change order by owner during construction	7	17	24	18	13	7	2.4
Delay in progress payment by owner	5	2	15	27	26	11	3.16
Poor communication and coordination by contractor with other parties	7	18	23	23	12	3	2.28
Difficulties in financing project by contractor	6	9	12	31	21	7	2.85
Delay in material delivery	13	21	20	19	11	2	2
Delay in obtaining permits from regulatory authorities	39	16	8	10	10	3	1.36
Poor site management and supervision by contractors	8	22	22	19	11	4	2.17
Poor communication and coordination by owner with other parties	17	24	23	12	8	2	1.72
Delay in approving shop drawing and sample materials	17	20	25	11	10	3	1.84
Inadequate definition of substantial completion	19	23	18	15	7	4	1.77
Delay in approving major changes in the scope of work by consultant	7	24	25	14	11	5	2.15

Frequent change of sub-contractors because of their inefficient work	17	37	19	8	4	1	1.4
Poor communication/coordination between consultant & other parties	8	20	26	16	11	5	2.2
Rain effect on construction activities	21	16	16	21	9	3	1.88
Conflicts between contractor and other parties (consultant and owner)	17	18	19	15	12	5	2.02
Delay in delivering the site to the contractor by the owner	25	27	6	14	11	3	1.63
Slowness in decision making process by the owner	21	11	24	15	13	2	1.93
Shortage of equipment	25	24	12	17	7	1	1.53
Shortage of labourers	29	15	14	10	12	6	1.76
Unclear and inadequate drawings	17	24	22	11	8	4	1.78
Equipment breakdowns	22	30	14	15	5	0	1.43
Ineffective delay penalties	27	19	13	11	11	5	1.71
Changes in government regulations/laws	23	30	13	12	7	1	1.45
Inadequate experience of consultant	27	23	7	16	11	2	1.62
Low level of equipment operator's skill	30	28	16	3	7	2	1.24
Insufficient data collection and survey before design	6	17	17	19	16	11	2.64
Delays in producing design documents	11	13	23	19	14	6	2.35
Poor qualification of the contractor's technical staff	17	16	11	26	10	6	2.16
Mistakes and discrepancies in design documents	15	28	18	16	7	2	1.74
Inflexibility (rigidity) of consultant	13	32	18	17	5	1	1.67
Complexity of project design	31	8	18	12	9	8	1.81
Lateness in selection of finishing materials	16	20	21	15	9	5	1.95
Low productivity of labourers	30	20	21	8	5	2	1.35
Suspension of work by owner	31	9	9	15	16	6	1.93
Lack of high technology mechanical equipment	32	19	13	9	10	3	1.48
Delay in performing inspection & testing by consultant	12	23	20	17	11	3	2.01
Delay in providing services from utilities (such as water, electricity, e.t.c)	20	14	19	14	14	5	2.03
Inadequate design team experience	13	31	17	11	11	3	1.83

Source: Field work (2013)

4.5 Multivariate Model Development

4.5.1 Principal component analysis of the qualitative factors

Taking into consideration, Field's (2005) observation that; when questions are designed to measure the same underlying dimension (or dimensions), it should be expected that they will correlate with each other (since they are measuring the same thing), it was anticipated that the qualitative factors considered in this study would be closely related, since they were all used to address the same question in the questionnaire survey. A principal component analysis was therefore performed on these qualitative factors in order to identify a reduced set of uncorrelated principal components which can summarize the qualitative factors without much loss of information in the process and

thus, prevent the likely occurrence of multicolinearity. The results of the PCA were as shown in Tables; 11, 12, 13 and 14.

4.5.1.1 Testing for the appropriateness of using a PCA

Table 11: The KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.732
Bartlett's Test of Sphericity	Approx. Chi-Square	7.760E3
	df	990
	Sig.	.000

Table 11 above, shows the Kaiser-Meyer-Olkin measure of sampling adequacy and the Bartlett's test of sphericity, both of which are tests carried out to determine the appropriateness of using a PCA. The KMO varies between 0 and 1, with; values between 0.5 and 0.7 signifying that the sampling adequacy is mediocre; values between 0.7 and 0.8 signifying that the sampling adequacy is good; values between 0.8 and 0.9 signifying that the sampling adequacy is great; values above 0.9 signifying that the sampling adequacy is superb and values below 0.5 indicating the need for more data to be collected (Field, 2005). The Bartlett's test on the other hand tests the null hypothesis that the original correlation matrix is an identity matrix and therefore, for this not to be true, the test must always be significant at $p < 0.05$.

As shown in Table 11, the KMO had a value of 0.732, which fall within the range of 0.7 to 0.8, indicating that the sampling adequacy for this study was good. Also, as shown in the table, the Bartlett's test was highly significant ($P < 0.05$), indicating that the original correlation matrix was not an identity matrix and also confirming the speculation that the qualitative factors were closely related, to be true. Both tests therefore confirmed that performing PCA on the qualitative factors was appropriate.

4.5.1.2 Extracting principal components

Table 12: Extraction of Principal Components

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	22.442	49.87	49.87	22.442	49.87	49.87	9.701	0.558	21.558
2	4.42	9.823	59.693	4.42	9.823	59.693	7.57	2116.82	38.381
3	2.884	6.408	66.101	2.884	6.408	66.101	5.19	11.532	49.913
4	2.55	5.667	71.769	2.55	5.667	71.769	4.218	9.374	59.287
5	1.958	4.351	76.119	1.958	4.351	76.119	3.556	7.902	67.19
6	1.759	3.909	80.028	1.759	3.909	80.028	2.56	5.688	72.878
7	1.426	3.168	83.196	1.426	3.168	83.196	2.46	5.468	78.345
8	1.222	2.717	85.913	1.222	2.717	85.913	2.066	4.592	82.937
9	1.127	2.505	88.418	1.127	2.505	88.418	1.868	4.151	87.088
10	1.026	2.281	90.699	1.026	2.281	90.699	1.625	3.611	90.699
11	0.782	1.738	92.438						
12	0.602	1.339	93.776						
13	0.559	1.242	95.018						
14	0.48	1.067	96.085						
15	0.365	0.81	96.895						
16	0.217	0.483	97.378						
17	0.19	0.422	97.8						
18	0.144	0.319	98.119						
19	0.132	0.293	98.412						
20	0.122	0.272	98.684						
21	0.067	0.148	98.832						
22	0.062	0.137	98.969						
23	0.061	0.136	99.105						
24	0.057	0.126	99.232						
25	0.051	0.112	99.344						
26	0.043	0.095	99.439						
27	0.035	0.078	99.517						
28	0.031	0.068	99.586						
29	0.029	0.065	99.651						
30	0.027	0.061	99.712						
31	0.025	0.056	99.768						
32	0.021	0.046	99.814						
33	0.013	0.03	99.843						
34	0.013	0.029	99.872						
35	0.011	0.024	99.896						
36	0.01	0.022	99.918						
37	0.008	0.019	99.936						
38	0.007	0.015	99.951						
39	0.006	0.014	99.965						
40	0.005	0.01	99.975						
41	0.004	0.008	99.983						
42	0.003	0.007	99.991						
43	0.002	0.005	99.995						
44	0.002	0.003	99.999						
45	0.001	0.001	100						

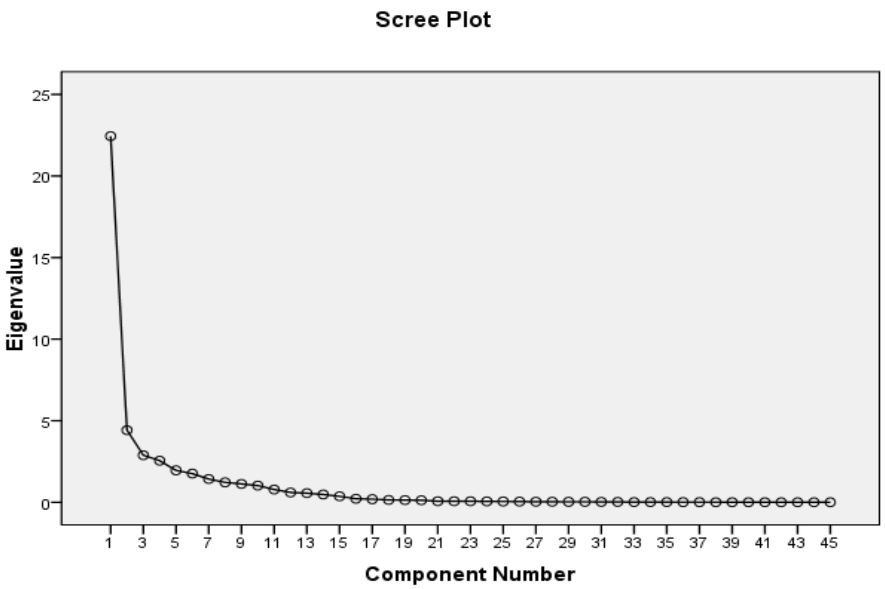


Figure 2: The Scree Plot

Table 12 (labelled extraction of principal components) and Figure 2, both show the process of extracting the principal components in a principal component analysis. Table 12 shows the eigenvalues associated with each linear component before extraction, after extraction and after rotation while Figure 2 (the scree plot) shows a plot of the eigenvalues against the linear components. The eigenvalue associated with each linear component represent the variance explained by that linear component and as recommended by Field (2005), only eigenvalues greater than 1 are considered when extracting the principal components.

As shown in Table 12, 45 linear components were generated for this study basically because 45 qualitative factors were considered in the study. However, looking through the eigenvalues associated with each linear component, it can be seen that there are only 10 linear components with eigenvalues greater than 1 and therefore there will be only 10

principal components as shown in the column labelled extraction sum of squared loadings. As shown in this Table also, the total variance explained/accounted for by the extracted principal components, was 90.699%, indicating that very little information (9.30%) was lost in the process and therefore good results will still be obtained when these principal components are used, instead of the whole data set. The first principal component accounted for the highest variance (49.87%), followed by the second component (9.823%). This was however not the case in the column labelled rotation sums of squared loadings, as rotation in PCA is meant to equalize the relative importance of each principal component.

As for the case of Figure 2 (the scree plot), the point of inflection at which the plot begins to tail off, signifies the point at which the linear components with eigenvalues greater than 1 terminates (Field, 2005). Taking a close look at this plot, it can be seen that ten principal components were also indicated by the plot, although this was not as obvious as the case of searching for eigenvalues greater than 1.

4.5.1.3 Selecting the principal corresponding components

Table 13: Rotated Component Matrix

	Component										
	1	2	3	4	5	6	7	8	9	10	h^2
Delay in progress payment by owner	0.93										0.94
Complexity of project design	0.84										0.95
Unclear and inadequate drawings	0.79										0.94
Shortage of labourers	0.76										0.95
Frequent change of sub-contractors because of their inefficient work	0.72										0.88
Poor communication and coordination by owner with other parties	0.70										0.88
Low level of equipment operators skill	0.70										0.91
Equipment breakdowns	0.67										0.92
Inadequate definition of substantial completion	0.67										0.79
Low productivity of labourers	0.65										0.93
Inadequate design team experience	0.65										0.94

Delay in performing inspection & testing by consultant	0.64								0.91
Rework due to errors during construction	0.57								0.93
Delay in providing services from utilities (such as water, electricity, e.t.c)	0.54	0.54							0.94
Lateness in selection of finishing materials		0.81							0.95
Improper construction methods implemented by contractor		0.75							0.89
Conflicts in sub-contractors schedule in execution of project		0.73							0.93
Late procurement of materials		0.70							0.86
Poor communication and coordination by contractor with other parties		0.66							0.88
Poor qualification of the contractors technical staff		0.65							0.92
Delay in approving shop drawing and sample materials		0.64							0.91
Difficulties in financing project by contractor		0.64							0.74
Ineffective planning and scheduling of project by contractor		0.60							0.87
Conflicts between contractor and other parties (consultant and owner)		0.55							0.95
Lateness in revising and approving design document by owner			0.84						0.85
Rain effect on construction activities			0.78						0.82
Slowness in decision making process by the owner			0.67						0.87
Delay in approving major changes in the scope of work by consultant			0.67						0.94
Inflexibility (rigidity) of consultant									0.91
Change order by owner during construction				0.90					0.89
Changes in government regulations/laws				0.76					0.94
Lack of high technology mechanical equipment				0.61					0.96
Inadequate experience of consultant				0.55	0.51				0.95
Suspension of work by owner					0.73				0.96
Ineffective delay penalties					0.68				0.93
Mistakes and discrepancies in design documents									0.94
Shortage of equipment									0.91
Insufficient data collection and survey before design						0.76			0.81
Delay in material delivery						0.70			0.92
Delays in producing design documents			0.50			0.58			0.91
Delay in delivering the site to the contractor by the owner							0.85		0.95
Poor site management and supervision by contractors								0.80	0.92
Delay in obtaining permits from regulatory authorities								-0.71	0.82
Delay in sub-contractors work								0.69	0.94
Poor communication/coordination between consultant & other parties									0.58 0.90

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

In order to be able to use the extracted principal components in a multiple linear regression with the quantitative factors, it was necessary to select the principal

corresponding component/the real signification of the extracted principal components. The criterion for selection used in the research work of Dascalu and Cozma (2007) was adopted for this purpose. This criterion involves selecting the predictor (variable/linear component) with the biggest correlation coefficient in the rotated component matrix. In addition to this, when two predictors have very close correlation coefficient, the principal component is the predictor with smaller correlation coefficients with the other principal components from the matrix.

Table 13 shows the rotated component matrix for the principal components extracted in Table 12. It shows the factor loadings of each variable/linear component into the extracted principal components. Factor loading refer to the correlation coefficient between each variable/linear component and a principal component. As shown in Table 13, each variable/linear component loaded differently into the respective principal components. This clearly indicates that while the correlation coefficient/factor loading of a variable/linear component may be highly significant in one of the principal components, it may not be significant in the other principal components and hence, the omission/absence of its correlation coefficient/factor loading in such principal components.

Going by the selection criterion stated earlier, it can be seen from Table 13 that the variables/linear components with the highest correlation coefficient and thus the principal corresponding components were as shown in table 14.

Table 14: Selected Principal Corresponding Components

Principal corresponding component		Factor loading
1	Delay in progress payment by owner	0.93
2	Change order by owner during construction	0.90
3	Delay in delivering the site to the contractor by the owner	0.85
4	Complexity of project design	0.84
5	Lateness in revising and approving design document by owner	0.84
6	Lateness in selection of finishing materials	0.81
7	Poor site management and supervision by contractors	0.80
8	Unclear and inadequate drawings	0.79
9	Rain effect on construction activities	0.78
10	Insufficient data collection and survey before design	0.76

4.5.2 Multiple linear regression analysis to determine the nature of the relationship between construction time and construction time influencing factors

Based on an assumption that a multiple linear relationship of the form; $y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots\ldots + \beta_q x_{qi} + \varepsilon_i$, existed between construction duration and a combination of the quantitative and qualitative factors, a multiple linear regression analysis was performed using construction duration as the dependent/response variable, and the three quantitative factors (cost, GFA, and number of floors) as well as the ten principal components of the qualitative factors, as the independent/explanatory variables. The result of the multiple regression analysis which was performed using a forward selection procedure at a significance level of 5%, were as shown in Tables 15, 16 and 17.

Table 15: Summary of Fitted Regression Model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.680[a]	.462	.455	244.643	.462	72.071	1	85	.000
2	.711[b]	.506	.494	235.888	.044	7.351	1	84	.008
3	.743[c]	.553	.536	225.716	.047	8.649	1	83	.004
4	.772[d]	.595	.575	216.037	.043	8.513	1	82	.005

a. Predictors: (Constant), Estimated final Cost (in millions of Naira)

b. Predictors: (Constant), Estimated final Cost (in millions of Naira), Insufficient data collection and survey before design

c. Predictors: (Constant), Estimated final Cost (in millions of Naira), Insufficient data collection and survey before design, GFA (m2)

d. Predictors: (Constant), Estimated final Cost (in millions of Naira), Insufficient data collection and survey before design, GFA (m2), Delay in progress payment by owner

e. Dependent Variable: Duration (in days)

Table 16: ANOVA Results of Regression Analysis

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	4313483	1	4313483	72.071	.000[a]
	Residual	5027407	85	59850.08		
	Total	9340890	86			
2	Regression	4722527	2	2361263	42.436	.000[b]
	Residual	4618363	84	55642.93		
	Total	9340890	86			
3	Regression	5163176	3	1721059	33.781	.000[c]
	Residual	4177714	83	50947.73		
	Total	9340890	86			
4	Regression	5560472	4	1390118	29.785	.000[d]
	Residual	3780418	82	46671.83		
	Total	9340890	86			

Table 17: Regression Coefficients of Fitted Model

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
4	(Constant)	177.175	61.436		2.884	0.005	54.936	299.413		
	Estimated final Cost (in millions of Naira)	9.25E-08	0	0.614	8.334	0	0	0	0.92	1.087
	Insufficient data collection and survey before design	65.903	17.495	0.292	3.767	0	31.093	100.714	0.831	1.204
	GFA (m2)	0.078	0.019	0.372	4.135	0	0.04	0.115	0.616	1.622
	Delay in progress payment by owner	52.599	18.028	0.251	2.918	0.005	88.469	16.729	0.677	1.477

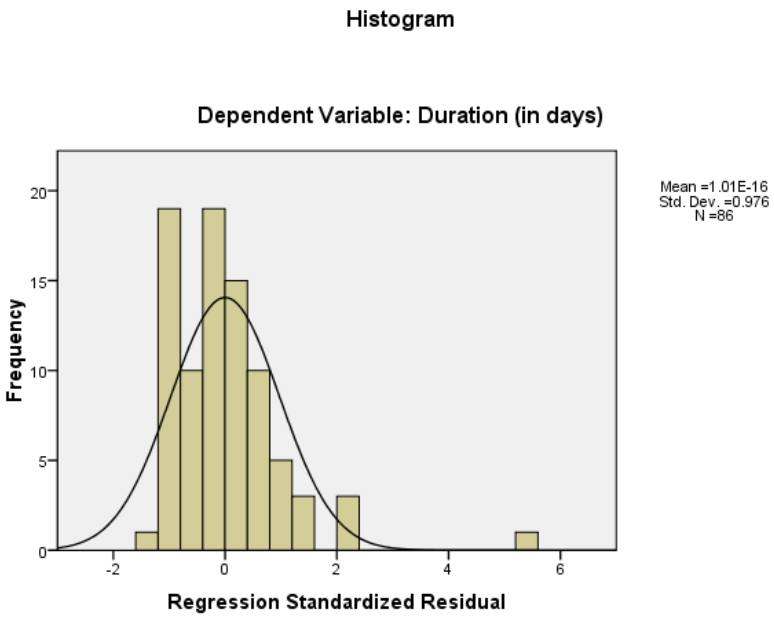


Figure 3: Error Distribution of Fitted Model

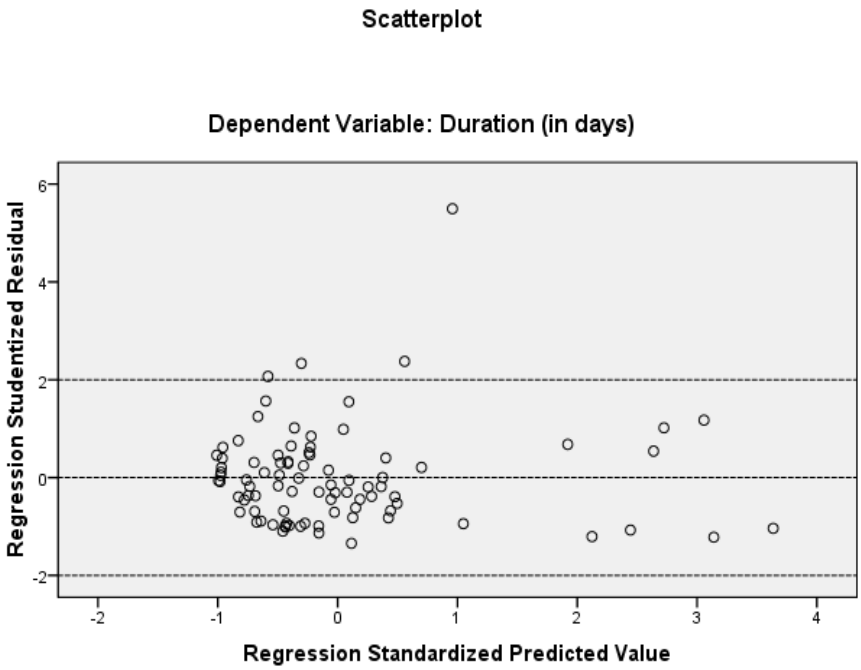


Figure 4: Homogeneity of Errors of Fitted Model

4.5.2.1 Discussion of regression results

As shown in Table 15, the forward selection process was iterated through five steps. In the first step, the quantitative variable estimated final cost was selected as it explained 46.2% of the total variation in the response variable (duration (in days)), with the test for the single regression coefficient also shown to be highly significant at ($F (1, 85) = 72.071$, $P < 0.05$). In the second step, the qualitative variable insufficient data collection and survey before design was included as it increased the percentage variance explained (R^2) by 4.4% at a significance level of ($F (1, 84) = 7.351$, $P < 0.05$). In the third step, GFA increased the percentage variance explained by 4.7% at a significance level of ($F (1, 83) = 8.649$, $P < 0.05$), and so it was added to the model. In the fourth step, the highest increase in the percentage variance explained was from the qualitative variable; delay in progress payment by owner (4.3%), at a significance level of ($F (1, 82) = 8.513$, $P < 0.05$). Delay in progress payment by the owner was therefore added to the model in

the fourth step. In the fifth step, no variable was added as the remaining potential predictor variables did not improve the model according to the chosen inclusion criterion of an increase in R^2 at a 5% significance level.

The final fitted model (model 4), as shown in Table 15, comprised four predictors; estimated final cost, insufficient data collection and survey before design, GFA and delay in progress payment by the owner. The multiple correlation coefficient was close to 1 ($R = 0.77$), indicating that there is a strong correlation between the observed duration times and those predicted by the fitted model. Also, as shown in Table 16, the fitted model was highly significant at the 5% significance level ($F(4, 82) = 29.785$, $P < 0.001$).

However, as indicated by the fitted model's coefficient of determination (R^2), only 59.5% of the total variation in duration was explained by the model, implying that as much as 40.5% of the variability in duration will not be accounted for when the model is used to make predictions. Similarly, as shown Figures 3 and 4, apart from the fact that the error distribution of the model is not normally distributed, the fitted model appears to have been influenced by an outlier, which most likely may be responsible for the very small confidence interval of some of the model's predictors, as shown in Table 17.

Based on the foregoing, it was necessary to; eliminate the identified outlier and; transform the study's data set in order to normalize the error distribution of the model and improve the confidence interval of the predictor's coefficients. It was also necessary to; in line with Kaka and Price (1993)'s observation that poor classification of projects has been indicated to be one of the main reasons for the unsuccessful results of some previous models, partition the study's data set in order to check if the low R^2 of the model was as a result of a poor classification.

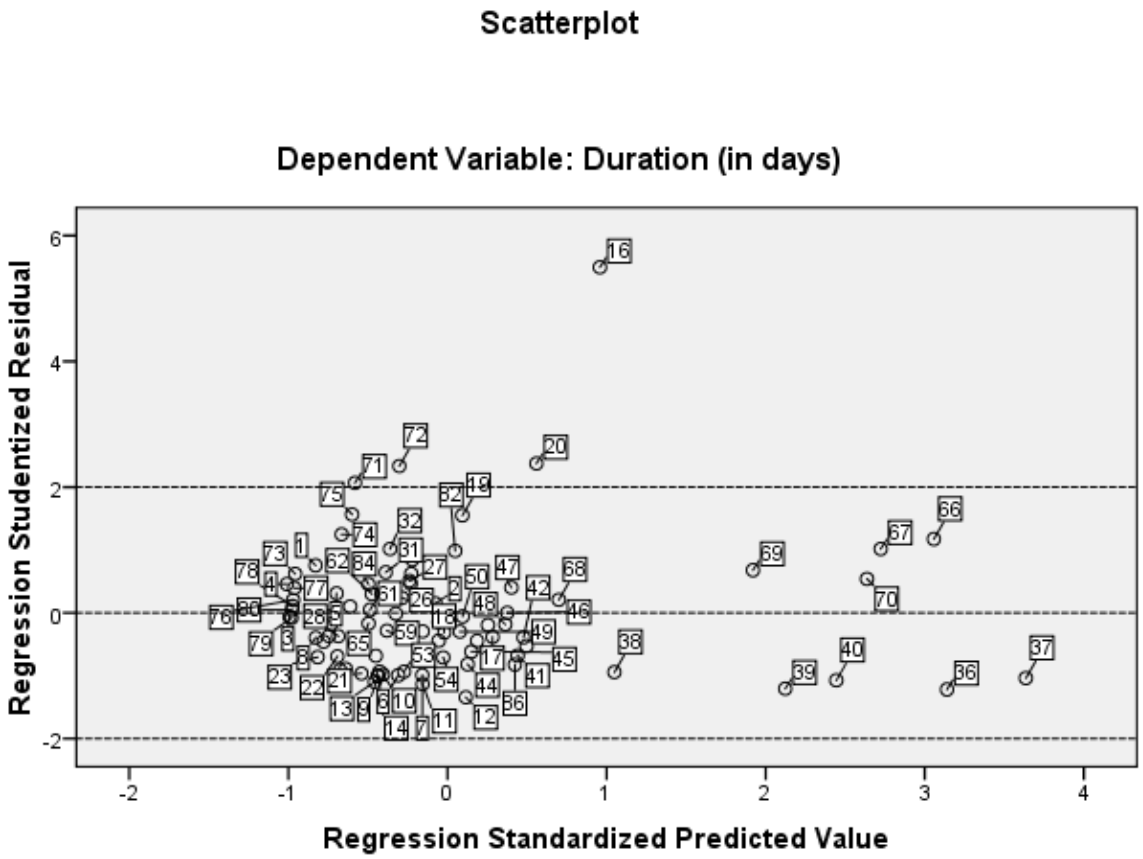


Figure 5: Identification of outliers

As shown in Figure 5, a total of 3 observations (observations; 16, 20 and 72) were identified as influential outliers and were therefore eliminated. The double log transformation was used to transform the data set which was then partitioned/re-classified into data set corresponding to public and private sector projects. Because a likert scale which ranged from zero to five was used in the survey questionnaire, the data set for the study thus, contained zeros. Hence, as log transformation can only be applied to numbers above zero, it was necessary to add to each number in the data set, a constant number such that each number in the data set were all above zero, before the log transformation was applied. The results of the multiple linear regression analysis (using a

forward selection procedure at a significance level of 5%) which was repeated following these adjustments were as shown below.

4.5.2.2 Results after transformation and partitioning of the study's data set

Regression results (all projects):

Table 18: Summary of Fitted Regression Model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.557[a]	.311	.302	.54785	.311	36.935	1	82	.000
2	.666[b]	.443	.429	.49543	.133	19.273	1	81	.000
3	.689[c]	.475	.455	.48412	.032	4.827	1	80	.031
4	.719[d]	.518	.493	.46685	.043	7.028	1	79	.010

a. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira)

b. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner

c. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner, Ln Poor site management and supervision by contractors

d. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner, Ln Poor site management and supervision by contractors, Ln GFA (m2)

e. Dependent Variable: Ln Duration (in days)

Table 19: ANOVA Results of Regression Analysis

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	11.086	1	11.086	36.935	.000[a]
	Residual	24.612	82	.300		
	Total	35.698	83			
2	Regression	15.816	2	7.908	32.219	.000[b]
	Residual	19.881	81	.245		
	Total	35.698	83			
3	Regression	16.948	3	5.649	24.103	.000[c]
	Residual	18.750	80	.234		
	Total	35.698	83			
4	Regression	18.479	4	4.620	21.196	.000[d]
	Residual	17.218	79	.218		
	Total	35.698	83			

Table 20: Regression Coefficients of Fitted Model

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
4	(Constant)	3.961	0.567		3.122	0.003	0.642	2.9		
	Ln Estimated final Cost (in millions of Naira)	3.191	0.035	3.529	5.435	0	0.121	1.261	0.645	1.55
	Ln Delay in progress payment by owner	2.856	0.198	2.404	3.784	0	1.538	2.75	0.576	1.737
	Ln Poor site management and supervision by contractors	3.687	0.232	3.265	2.968	0.004	0.226	1.149	0.766	1.305
	Ln GFA (m2)	3.172	0.065	3.294	2.651	0.01	0.043	1.301	0.495	2.021

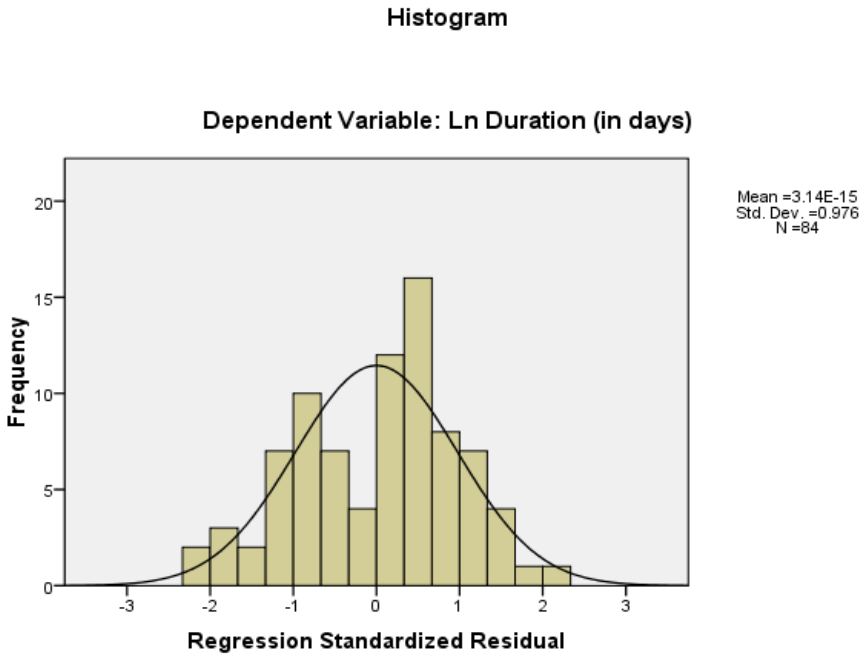


Figure 6: Error Distribution of Fitted Model

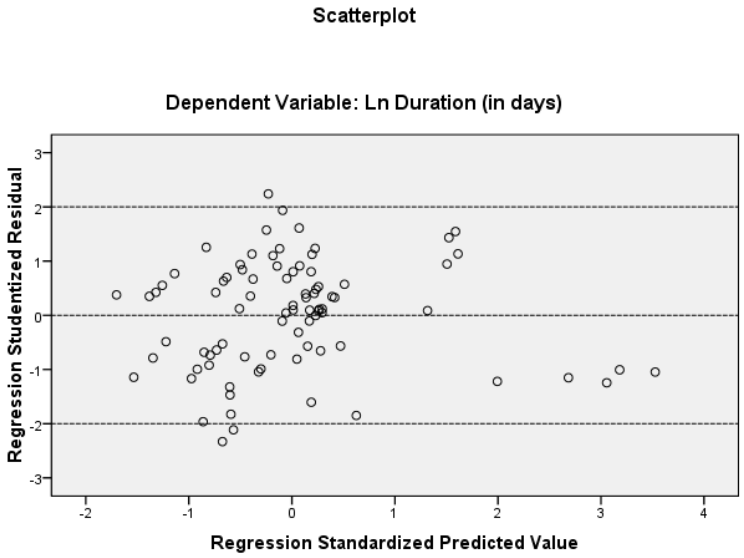


Figure 7: Homogeneity of Errors of Fitted Model

As shown in Table 18, after the identified outliers were eliminated and the study's data set transformed (but not partitioned), the final fitted model still comprised four

predictors; estimated final cost, delay in progress payment by the owner, poor site management and supervision by contractors and GFA. The multiple correlation coefficient (R) decreased (from 0.772 to 0.719) but still remained close to 1, indicating that the correlation between the observed duration times and those predicted by the fitted model remained very strong. The fitted model also remained statistically significant as shown in Table 19 ((F (4, 79) = 21.196, P < 0.001), while the error distribution of the model clearly became normal as shown in Figure 6.

However, as shown in Table 18, the amount of variability in the dependent variable (duration) explained by the model still remained very low (0.518), indicating that as much as 48.20% of the variability in duration will not be accounted for when the model is used to make predictions.

Regression results (public sector projects):

Table 21: Summary of Fitted Regression Model

						Change Statistics			
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.667[a]	0.445	0.434	0.545	0.445	40.915	1	52	0
2	.730[b]	0.533	0.514	0.505	0.088	9.383	1	51	0.004
3	.808[c]	0.653	0.632	0.44	0.12	17.003	1	50	0
4	.872[d]	0.761	0.741	0.369	0.108	21.62	1	49	0
5	.910[e]	0.829	0.811	0.315	0.068	18.642	1	48	0
6	.945[f]	0.892	0.878	0.253	0.063	27.088	1	47	0

a. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira)

b. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner

c. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner, Ln GFA (m2)

d. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner, Ln GFA (m2), Ln Lateness in revising and approving design document by owner

e. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner, Ln GFA (m2), Ln Lateness in revising and approving design document by owner, Ln Delay in delivering the site to the contractor by the owner

f. Predictors: (Constant), Ln Estimated final Cost (in millions of Naira), Ln Delay in progress payment by owner, Ln GFA (m2), Ln Lateness in revising and approving design document by owner, Ln Delay in delivering the site to the contractor by the owner, Ln Change order by owner during construction

g. Dependent Variable: Ln Duration (in days)

Table 22: ANOVA Results of Regression Analysis

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12.159	1	12.159	40.915	.000[a]
	Residual	15.444	52	0.297		
	Total	27.603	53			
2	Regression	14.554	2	7.277	28.512	.000[b]
	Residual	13.005	51	0.255		
	Total	27.559	53			
3	Regression	17.841	3	5.947	30.759	.000[c]
	Residual	9.650	50	0.193		
	Total	27.491	53			
4	Regression	20.783	4	5.196	38.182	.000[d]
	Residual	6.664	49	0.136		
	Total	27.447	53			
5	Regression	22.638	5	4.528	45.501	.000[e]
	Residual	4.800	48	0.1		
	Total	27.438	53			
6	Regression	24.371	6	4.062	63.479	.000[f]
	Residual	3.008	47	0.064		
	Total	27.379	53			

Table 23: Regression Coefficients of Fitted Regression Model

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
6	(Constant)	4.915	0.556		4.418	0	1.337	3.577		
	Ln Estimated final Cost (in millions of Naira)	2.051	0.027	2.146	1.899	0.004	-0.003	0.104	0.398	2.51
	Ln Delay in progress payment by owner	0.442	0.143	1.23	10.931	0	1.845	1.271	0.472	2.117
	Ln GFA (m2)	2.646	0.055	3.015	11.634	0	0.534	0.757	0.308	3.246
	Ln Lateness in revising and approving design document by owner	3.659	0.172	3.597	9.66	0	1.313	2.004	0.613	1.632
	Ln Delay in delivering the site to the contractor by the owner	1.134	0.16	1.63	5.404	0	1.188	0.543	0.499	2.005
	Ln Change order by owner during construction	1.185	0.157	1.692	5.205	0	1.13	0.5	0.671	1.491

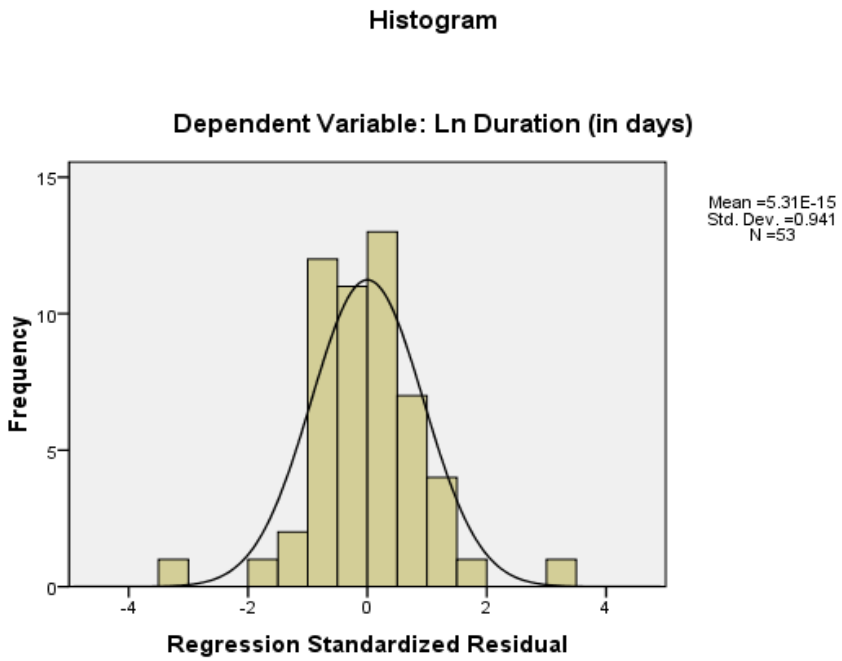


Figure 8: Error Distribution of Fitted Regression Model

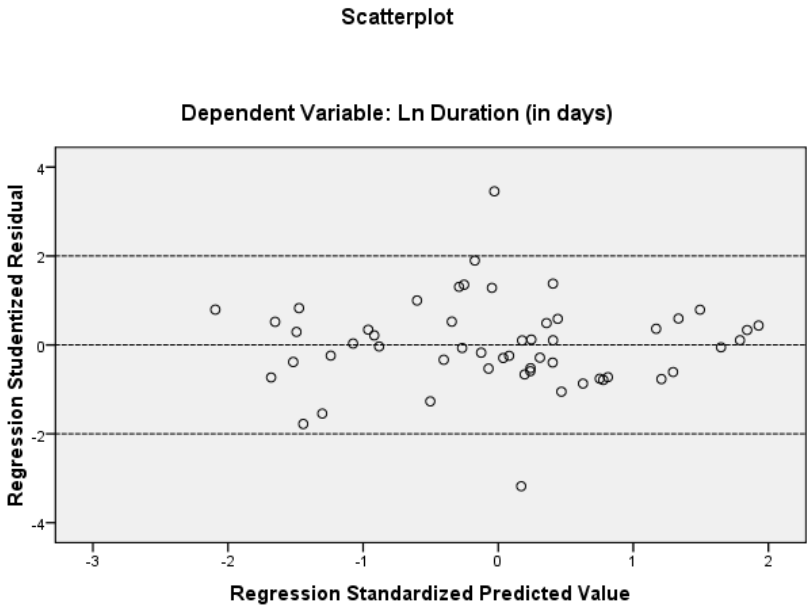


Figure 9: Homogeneity of Errors of Fitted Regression Model

As shown in Table 21, the final fitted model for the data set corresponding to public sector projects, comprised 6 predictors; estimated final cost (in millions of Naira), delay

in progress payment by owner, GFA (m2), lateness in revising and approving design document by owner, delay in delivering the site to the contractor by the owner, and change order by owner during construction. The multiple correlation coefficient (R) and the coefficient of determination (R^2) are both very close to 1 (0.945 and 0.892 respectively), indicating that the observed durations and the fitted model's predictions are strongly correlated and also, a very large proportion of the variability in duration (89.2%) will be accounted for when the model is used to make predictions. The f-statistics as shown in Table 22, indicates that the fitted model is highly significant ((F (6, 47) = 63.479, P < 0.001), while the variance inflation factors (VIFs) as shown in Table 23, are all bellow 5, indicating the non-existence of multicolinearity between the independent variables.

Regression results (private sector projects):

Table 24: Summary of Fitted Regression Model

						Change Statistics			
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.529[a]	0.28	0.254	0.437	0.28	10.878	1	29	0.003
2	.811[b]	0.658	0.633	0.307	0.378	29.905	1	28	0
3	.879[c]	0.773	0.747	0.255	0.114	13.101	1	27	0.001
4	.900[d]	0.811	0.781	0.237	0.038	5.035	1	26	0.034

a. Predictors: (Constant), Ln Complexity of project design

b. Predictors: (Constant), Ln Complexity of project design, Ln Poor site management and supervision by contractors

c. Predictors: (Constant), Ln Complexity of project design, Ln Poor site management and supervision by contractors, Ln Estimated final Cost (in millions of Naira)

d. Predictors: (Constant), Ln Complexity of project design, Ln Poor site management and supervision by contractors, Ln Estimated final Cost (in millions of Naira), Ln Rain effect on construction activities

e. Dependent Variable: Ln Duration (in days)

Table 25: ANOVA Results of Regression Analysis

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	2.080	1	2.080	10.878	.003[a]
	Residual	5.539	29	.191		
	Total	7.619	30			
2	Regression	4.894	2	2.447	26.006	.000[b]
	Residual	2.632	28	.094		
	Total	7.526	30			
3	Regression	5.745	3	1.915	29.475	.000[c]
	Residual	1.755	27	.065		
	Total	7.500	30			
4	Regression	6.028	4	1.507	26.795	.000[d]
	Residual	1.462	26	.056		
	Total	7.490	30			

Table 26: Regression Coefficients of Fitted Regression Model

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
4	(Constant)	4.574	1.247		0.053	0.002	1.634	2.501		
	Ln Complexity of project design	4.527	0.154	4.964	9.586	0	1.789	2.156	0.648	1.544
	Ln Poor site management and supervision by contractors	4.511	0.238	3.626	6.362	0	1.022	3	0.78	1.282
	Ln Estimated final Cost (in millions of Naira)	3.16	0.044	3.357	3.65	0.001	1.07	2.25	0.789	1.268
	Ln Rain effect on construction activities	4.497	0.667	3.219	2.244	0.034	0.123	2.872	0.792	1.262

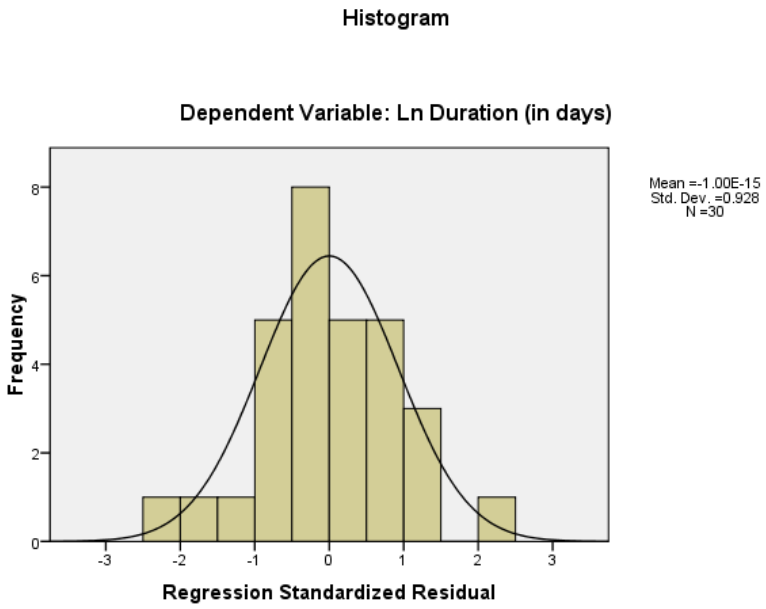


Figure 10: Error Distribution of Fitted Regression Model

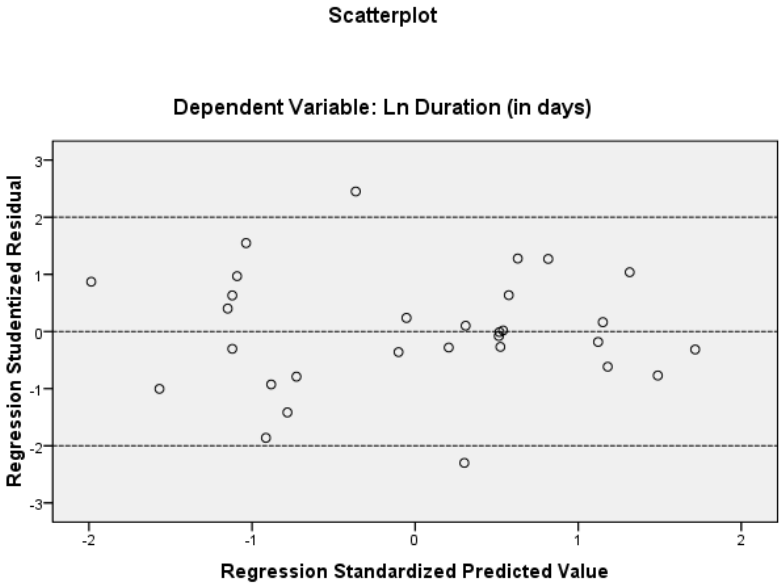


Figure 11: Homogeneity of Errors of Fitted Regression Model

The final fitted model for the data set corresponding to private sector projects as shown in Table 24, comprised four predictors; complexity of project design, poor site management and supervision by contractors, estimated final cost (in millions of Naira),

and rain effect on construction activities. Also, just like the public sector projects, the multiple correlation coefficient (R) and the coefficient of determination (R^2) are both very close to 1 (0.900 and 0.811 respectively), indicating that the observed durations and the fitted model's predictions are strongly correlated and also, a very large proportion of the variability in duration (81.1%) will be accounted for when the model is used to make predictions. The fitted model is statistically significant as shown in Table 25 (($F(4, 26) = 26.795$, $P < 0.001$), while the variance inflation factors (VIFs) are all below 5 as shown in Table 26, indicating the non-existence of multicolineariy between the dependent variables.

4.5.2.3 Discussion of regression results following the transformation and partitioning of the study's data set

As shown in the regression results, after the identified outliers were eliminated and the data set transformed and partitioned to account for data corresponding to public and private sector projects, the amount of variability in duration explained by the model (for all projects) remained very small ($R^2 = 0.518$), while the amount of the variability in duration explained by the model for the public sector data as well as the private sector data were both very high ($R^2 = 0.892, 0.811$ respectively).

The improvement in the R^2 value (coefficient of determination) of the model, following the partitioning of the study's data set, clearly conforms to Kaka and Price (1993)'s assertion that poor classification of projects has been indicated to be one of the main reasons for the unsuccessful results of previous models. It shows that the initial classification considered for the study (building construction projects) was poor and invariably, responsible for the low R^2 value associated with the model, while the latter

classifications (public and private sector projects) were better classifications and thus, responsible for the improvement in the R^2 value of the model.

Also, as shown in the model summary of each model category, the R^2 change statistics reveals that the total variability accounted for by ln(cost) alone in each model category was 0.311, 0.445 and 0.114 respectively. This clearly conforms to earlier research works in Nigeria conducted by Ojo (2001) as well as Ogunsemi and Jagboro (2006), which both found the Bromilow (1969) time cost model to be inapplicable to the Nigerian construction industry. The regression results therefore shows that in line with Olupolola et al. (2010)'s study, several other factors influence construction time in Nigeria. For example, as shown in the regression results, the duration of public sector projects in Nigeria is influenced by; estimated final cost (in millions of Naira), delay in progress payment by owner, GFA (m2), lateness in revising and approving design document by owner, delay in delivering the site to the contractor by the owner, and change order by owner during construction. Similarly, as shown in the regression results, the duration of private sector projects is influenced by: complexity of project design, poor site management and supervision by contractors, estimated final cost (in millions of Naira), and rain effect on construction activities. The relative strengths of the influences of these variables been reflected in their respective coefficients in the regression models.

Furthermore, as shown in the regression coefficients of the fitted models, the qualitative factors which emerged as predictors of construction duration were: *delay in progress payment by owner, lateness in revising and approving design document by owner, delay in delivering the site to the contractor by the owner, change order by owner during construction, complexity of project design, poor site management and supervision by contractors, as well as rain effect on construction activities.* These are different from the qualitative factors which emerged as predictors of construction duration in existing

multivariate models developed by Chan and Chan (2002) as well as Hofman et al. (2006). This thus, clearly reinforces Nkado (1995)'s assertion that there is little consensus as to which combination of these more qualitative factors provides an accurate predictor of construction duration. It is also an indication that multivariate models are context specific, since no combination of qualitative factors are the same for all multivariate models.

Finally, as shown in the regression coefficients of the fitted models, Cost and GFA were the quantitative (project-scope) factors which emerged as predictors of construction duration. This is however, not unexpected as both scope factors have been shown by previous research works to have good predictive abilities. However, as shown in the model summary of each model category, the R^2 change statistics reveals that the total variability accounted for by the scope factors alone (ln(cost) and ln(GFA)), were: 0.35, 0.57 and 0.114 respectively. The remaining variability of 0.164, 0.322 and 0.697 respectively, were accounted for by the qualitative factors. It is clear therefore that the qualitative factors which emerged as predictors in each model category, improved the R^2 value of each model and thus, the accuracy of the models.

4.5.3 Developing the Models

Only when a model appears adequate in the light of the data should the fitted regression equation be interpreted (Landau and Everitt, 2004). As shown and discussed lately, the model for public sector projects as well as the model for private sector projects appeared adequate in the light of the data corresponding to each. The model for all projects also appeared adequate in the light of the data, except that it had a low R^2 value. Since each model seemed adequate in the light of its data, it was therefore important to carry on and develop these models.

However, before commencing the development of these models, it was important to understand how to transform back estimated parameters of interest and how to interpret the coefficients in a regression model obtained from a regression with log transformed variables.

4.5.3.1 Transforming back and interpreting coefficients

Transforming back log transformed parameters/variables:

Based on the properties of a log-normal distribution in terms of the original variable Y (i.e.; mean of variable Y = $e^{\mu+ \sigma^2/2}$; median or geometric mean of variable Y = e^{μ} and; variance of variable Y = $(e^{\sigma^2} - 1)\, e^{2\mu+\sigma^2}$), Yang (2012) contends that back transforming the predicted values of a log transformed response variable ln(Y) by exponentiation (e^{μ}), gives the predicted geometric mean of Y rather than the arithmetic mean of Y. According to Yang (2012), using the formula; Y = $e^{X\hat{\beta}+\hat{\sigma}^2/2}$ (where $\hat{\beta}$ and $\hat{\sigma}^2$ are the estimated regression coefficients and the mean square error (MSE), respectively) instead will give the predicted mean values of Y in the original scale. Taking a cue from the above argument by Yang (2012), the back transformation for the double log transformation performed in this study, was done using the formula Y = $e^{X\hat{\beta}+\hat{\sigma}^2/2}$ as prescribed by Yang (2012).

Interpreting coefficients of log transformed variables:

For a linear regression model given by $Y = \beta_0 + \beta_1 X_1 + \beta_n X_n$, the standard interpretation of the regression parameter β_n is that a one-unit change in the predictor results in β_n units change in the expected value of the response variable while holding all the other

predictors constant. However, as outlined by Yang (2012) for a log transformed variable such coefficients are routinely interpreted in terms of percent change.

In order to interpret and thus, develop and test the models developed in this study, the formulas derived by Yang (2012) for the interpretation in a linear regression setting when the dependent variable, or the independent variable, or both variables are log transformed, were adopted. These formulas are as shown in Table 27.

Table 27: Interpreting Linear Regressions with Log Transformed Variables

Variable transformed	Interpretation
Dependent variable only i.e $\ln(Y) = \beta_0 + \beta_1 X_1 + \beta_n X_n$	1 unit increase in X_n would result in $(e^{\wedge} \beta_n - 1)100$ percentage change in Y, while keeping all other variables constant
Independent variable only i.e $Y = \beta_0 + \beta_1 \ln X_1 + \beta_n \ln X_n$	1 percent change in X_n results in $\beta_1 \ln(1.01)$ change in Y, while keeping all other variables constant
Both dependent and independent variable i.e $\ln(Y) = \beta_0 + \beta_1 \ln X_1 + \beta_n \ln X_n$	1 percent change in X_n results in $100(1.01^{\wedge} \beta_n - 1)$ percent change in Y, while keeping all other variables constant (for β_n less than 10, $100(1.01^{\wedge} \beta_n - 1)$ can be approximated to β_n)

Source: Yang (2012)

4.5.3.2 Public sector model Development

A multiple linear relationship of the form $Y = \beta_0 + \beta_1 X_1 + \beta_n X_n$ was assumed during the modelling process. However, because the data was double log transformed, the relationship can be rewritten as $\ln(Y) = \hat{\beta}_0 + \hat{\beta}_1 \ln X_1 + \hat{\beta}_n \ln X_n$. Based on the coefficients for the public sector model shown in Table 23, the public sector model can be expressed

as; Ln(Y) = 4.915 + 2.051Ln(C) + 0.442Ln(Q_1) + 2.646Ln(G) + 3.659Ln(Q_2) + 1.134Ln(Q_3) + 1.185Ln(Q_4) …… (3)

Back transforming; the variables in equation 1 with the formula Y = $e^{X\hat{\beta}+\hat{\sigma}^2/2}$ (in order to obtain the predicted mean values of Y in the original scale) and; the coefficients with the formula $100(1.01\hat{}\ \beta_n - 1)$ (in order to obtain the percent change in Y resulting from 1 percent change in any of the variables), the public sector model can be further rewritten as:

Y = $(e^{4.915}e^{0.032})$ x $(e^{lnC^{0.021}}e^{0.032})$ x $(e^{lnQ_1{}^{0.004}}e^{0.032})$ x $(e^{lnG^{0.0265}}e^{0.032})$ x $(e^{lnQ_2{}^{0.037}}e^{0.032})$ x $(e^{lnQ_3{}^{0.0113}}e^{0.032})$ x $(e^{lnQ_4{}^{0.0119}}e^{0.032})$ …… (4)

By simplifying equation 4 further, the public sector model can be written as:

Y = 140.82 x $1.033C^{0.021}$ x $1.033Q_1{}^{0.004}$ x $1.033G^{0.0265}$ x $1.033Q_2{}^{0.037}$ x $1.033Q_3{}^{0.0113}$ x $1.033Q_4{}^{0.0119}$ …… (5)

Where:

Y = duration in days,

C = estimated final cost in millions,

G = GFA in m^2

Q_1 = dummy variable for delay in progress payment by owner

Q_2 = dummy variable for lateness in revising and approving design document by owner

Q_3 = dummy variable for delay in delivering the site to the owner by the contractor

Q_4 = change order by the owner during construction

When using this model (equation 3) to compute duration, 3 instead of 1 will be entered for the dummy variables Q_1, Q_2, Q_3 and Q_4. This is because as stated earlier in this chapter, because a likert scale which ranged from zero to five was used in the survey questionnaire, the data set for the study thus, contained zero. Hence, as log

transformation can only be applied to numbers above zero, it was necessary to add to each number in the data set, a constant number (3) such that each number in the data set were all above zero, before the log transformation was applied. For this reason, it is considered that entering 3 instead of 1 (for the dummy variables), will reflect the original scale used.

4.5.3.3 Private sector model Development

A multiple linear relationship of the form $Y = \beta_0 + \beta_1 X_1 + \beta_n X_n$ was assumed during the modelling process. However, because the data was double log transformed, the relationship can be rewritten as $\ln(Y) = \hat{\beta}_0 + \hat{\beta}_1 \ln X_1 + \hat{\beta}_n \ln X_n$. Based on the coefficients for the private sector model shown in Table 26, the private sector model can be expressed as; $\text{Ln(Y)} = 4.474 + 4.527\text{Ln}(Q_1) + 4.511\text{Ln}(Q_2) + 3.160\text{Ln(C)} + 4.497\text{Ln}(Q_3)$ …… (6)

Back transforming the variables and coefficients with the formulas; $Y = e^{X\hat{\beta}+\hat{\sigma}^2/2}$ (in order to obtain the predicted mean values of Y in the original scale) and; $100(1.01\hat{}\ \beta_n - 1)$ (in order to obtain the percent change in Y resulting from 1 percent change in any of the variables) respectively, the private sector model can be rewritten as:

$Y = (e^{4.574}e^{0.028}) \times (e^{lnQ_1{}^{0.04527}}e^{0.028}) \times (e^{lnQ_2{}^{0.0451}}e^{0.028}) \times (e^{lnC^{0.0316}}e^{0.028}) \times (e^{lnQ_3{}^{0.0450}}e^{0.028})$ …… (7)

This can be further simplified as:

$Y = 99.65 \times 1.028Q_1{}^{0.04527} \times 1.028Q_2{}^{0.0451} \times 1.028C^{0.0316} \times 1.028Q_3{}^{0.0450}$ ……. (8)

Where:

Y = duration in days

C = estimated final cost in millions

Q_1 = dummy variable for complexity of project design

Q_2 = dummy variable for poor site management and supervision by contractors

Q_3 = dummy variable for rain effect on construction activities

4.5.3.4 All projects model Development

A multiple linear relationship of the form $Y = \beta_0 + \beta_1 X_1 + \beta_n X_n$ was assumed during the modelling process. However, because the data was double log transformed, the relationship can be rewritten as $\ln(Y) = \hat{\beta}_0 + \hat{\beta}_1 \ln X_1 + \hat{\beta}_n \ln X_n$. Based on the coefficients for the all project model shown in Table 20, the all project model can be expressed as;

$\text{Ln(Y)} = 3.961 + 3.191\text{Ln(C)} + 2.856\text{Ln}(Q_1) + 3.687\text{Ln}(Q_2) + 3.172\text{Ln(G)}$ …… (9)

Back transforming the variables and coefficients with the formulas; $Y = e^{X\hat{\beta}+\hat{\sigma}^2/2}$ (in order to obtain the predicted mean values of Y in the original scale) and; $100(1.01^{\wedge}\beta_n - 1)$ (in order to obtain the percent change in Y resulting from 1 percent change in any of the variables) respectively, the all projects model can be rewritten as:

$Y = (e^{3.961}e^{0.109}) \text{ x } (e^{lnC^{0.0319}}e^{0.109}) \text{ x } (e^{lnQ_1{}^{0.0286}}e^{0.109}) \text{ x } (e^{lnQ_2{}^{0.0369}}e^{0.109}) \text{ x } (e^{lnG^{0.0317}}e^{0.109})$ …… (10)

This can be further simplified as:

$Y = 58.55 \text{ x } 1.115C^{0.0319} \text{ x } 1.115Q_1{}^{0.0286} \text{ x } 1.115Q_2{}^{0.0369} \text{ x } 1.115G^{0.0317}$ …… (11)

Where:

Y = duration in days

C = estimated final cost in millions

G = GFA in m^2

Q_1 = dummy variable for delay in progress payment by owner

Q_2 = dummy variable for poor site management and supervision by contractors.

4.6 Model testing and validation

In line with previous research works outlined in chapter two, models developed in this study were tested and validated by:

i. Checking the assumptions of linearity made in the modelling process i.e.: the errors have the same variance (homogeneity); the errors arise from a normal distribution; the relationship between each explanatory variable and the dependent variable is linear and; there is independence of the error (no serial correlation).

ii. Comparing the planner's estimates and the model's predictions for projects which were not used in the modelling process i.e.: 10% of the study's data set which was set aside during the modelling process.

4.6.1 Checking for the assumptions of linearity

As shown in the residual plots for the homogeneity of errors (Figures: 7, 9 and 11), it can be seen that for each model the homogeneity of variance assumption was not violated since in each of the residual plots, the residuals scatter randomly around the zero line and the degree of scatter appeared constant across the range of the predicted values. Similarly, the histogram of the residuals (for the distribution of errors) as shown in Figures 6, 8 and 10 were clearly mound shaped and symmetric with a mean of zero, indicating that for each model the errors arise from a normal distribution,. Finally, it can be seen that for each model; the partial plots of the raw residuals of duration against the respective predictors as shown in appendix A, reveals a linear relationship between each explanatory variable and the dependent variable, while; the variance inflation factors (VIFs) as shown in the coefficients Table of each model (Tables; 20, 23 and 26), clearly

indicate that there is independence of the errors (no serial correlation) as the VIFs were all below 5 in each of the models.

4.6.2 Comparing the planner's estimates and the model's predictions

Tables 28 to 33 shows the comparison of the planners' estimates with the model's predictions for projects not used in the model development process (10% of the study's data which was set aside during the model development process), together with the results of an independent t-test performed to check whether the two sets of predictions were significantly different from each other or not.

4.6.2.1 Comparing planner's estimates and the Public sector model's predictions

Table 28: Comparing Model Predictions to Planner's Estimates

Planned duration	Model prediction	%age difference
294	305	4
272	280	3
308	315	2
238	252	6
327	332	2

Table 29: Statistical Significance of the Difference between Model's Predictions and Planner's Estimates

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Duration	Equal variances assumed	0.017	0.9	-0.424	8	0.683	-8.8	20.747	-56.643	39.043
	Equal variances not assumed			-0.424	7.939	0.683	-8.8	20.747	-56.707	39.107

As shown in Table 28, the public sector model's predictions were higher than the planner's estimates for all five projects considered. However, as shown in the independent samples t-test for the significance of the difference between the two estimates, both the levene's test for equality of variance and the significance (2-tailed) value were greater than 0.05, indicating that the variability in the two sets of predictions is not significantly different (equal variance can be assumed) and there is no statistically significant difference between the means of the two sets of predictions respectively. These results justify the high R^2 value of the public sector model and shows that predictions made using this model will be reliable.

4.6.2.2 Comparing planner's estimates and the Private sector model's predictions

Table 30: Comparing Model Predictions to Planner's Estimates

Planned duration	Model prediction	%age difference
266	235	-12
336	287	-15
250	229	-8
273	224	-18
371	341	-8

Table 31: Statistical Significance of the Difference between Model's Predictions and Planner's Estimates

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Duration	Equal variances assumed	.050	.829	1.115	8	.297	36.000	32.280	-38.437	110.437
	Equal variances not assumed			1.115	7.993	.297	36.000	32.280	-38.448	110.448

As shown in table 30, the private sector model's predictions were lower than the planner's estimates for all five projects considered. However, as shown in the independent samples t-test for the significance of the difference between the two estimates, both the levene's test for equality of variance and the significance (2-tailed) value were greater than 0.05 (0.829 and 0.297 respectively), indicating that the variability in the two sets of predictions is not significantly different (equal variance can be assumed) and there is no statistically significant difference between the means of the two sets of predictions respectively. These results, just like the public sector model, justify the high R^2 value of the private sector model and shows that predictions made using this model will also be reliable.

4.6.2.2 Comparing planner's estimates and the all projects model's predictions

Table 32: Comparing Model Predictions to Planner's Estimates

Planned duration	Model prediction	%age difference
273	212	-22
266	242	-9
327	268	-18
238	186	-22
308	196	-36
273	208	-24
336	258	-23
250	174	-30
294	226	-23
371	252	-32

Table 33: Statistical Significance of the Difference between Model's Predictions and Planner's Estimates

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Duration	Equal variances assumed	0.609	.445	4.283	18	.000	71.400	16.671	36.375	106.425
	Equal variances not assumed			4.283	16.889	.001	71.400	16.671	36.209	106.591

As shown in Table 32, the all project model's predictions were far lower than the planner's estimates for all 10 projects considered. Although, as shown in the independent samples T-test for the significance of the difference between the two estimates, the levene's test for equality of variance was greater than 0.05 (0.129), indicating that the variability in the two sets of predictions is not significantly different (equal variance can be assumed), the significance (2-tailed) value was however, very low (0.000) indicating

clearly that there is a statistically significant difference between the means of the two sets of predictions. The the levene's test for equality of variance thus, contradicts the low R^2 value of the all projects model, while the significance (2-tailed) value justify the all projects model's low R^2 value. Based on these results, it can be seen that predictions made using the all projects model will not be reliable.

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary of findings

The findings of the study following the presentation, analysis and discussion of results carried out in chapter four, were as follows:

i. Ten(10), out of the several key qualitative factors influencing building construction time in Nigeria are uncorrelated principal components which can be used to summarize the entire qualitative factors influencing construction time in Nigeria without much loss of information. The ten factors are; delay in progress payment by owner, change order by owner during construction, delay in delivering the site to the contractor by the owner, complexity of project design, lateness in revising and approving design document by owner, lateness in selection of finishing materials, poor site management and supervision by contractors, unclear and inadequate drawings, rain effect on construction activities, and insufficient data collection and survey before design.

ii. The predictive ability of multivariate construction time predicting models in Nigeria is determined by the accuracy of the classification considered for the model; broad based classifications which aggregate groups of projects into a single category are poor classifications, while specific classifications which separate projects into their exact type are better classifications.

iii. Cost alone can only account for less than 50% of the total variation in the duration of building construction projects in Nigeria. The more greater percentage (over 50%) is accounted for by other project scope factors and qualitative/managerial factors which influence construction time in Nigeria, namely; GFA (m2), delay in progress payment by owner, lateness in revising and approving design document by owner, delay in delivering the site to the contractor by the owner, change order by owner during construction, complexity of project design, poor site management and supervision by contractors, as well as rain effect on construction activities.

iv. Quantitative (project scope related) factors and qualitative (management related) factors can be combined to predict the duration of public sector projects and private sector projects in Nigeria, in the forms shown below:

Public sector projects:

$Y = 140.82 \times 1.033^{C^{0.021}} \times 1.033^{Q_1^{0.004}} \times 1.033^{G^{0.0265}} \times 1.033^{Q_2^{0.037}} \times 1.033^{Q_3^{0.0113}} \times 1.033^{Q_4^{0.0119}}$

Where:

Y = duration in days,

C = estimated final cost in millions of Naira,

G = GFA in m^2

Q_1 = dummy variable for delay in progress payment by owner

Q_2 = dummy variable for lateness in revising and approving design document by owner

Q_3 = dummy variable for delay in delivering the site to the owner by the contractor

Q_4 = change order by the owner during construction

Private sector projects:

$Y = 99.65 \times 1.028^{Q_1^{0.04527}} \times 1.028^{Q_2^{0.0451}} \times 1.028^{C^{0.0316}} \times 1.028^{Q_3^{0.0450}}$

Where:

Y = duration in days

C = estimated final cost in millions of Naira

Q_1 = dummy variable for complexity of project design

Q_2 = dummy variable for poor site management and supervision by contractors

Q_3 = dummy variable for rain effect on construction activities

Each of the forms/models stated above have been tested and validated in the study. In both forms/models, three (3) instead of one (1) is to be used for the dummy variables. This is because; a likert scale which ranged from zero to five was used in the study's survey questionnaire and as log transformation can only be applied to numbers above zero, it was therefore necessary to add a constant number (3 in the case of this study) to each number in the study's data set, in order to eliminate all zeros obtained in the data set (as a result of the 0 to 5 likert scale used), before the log transformation was applied. For this reason, it is considered that using 3 instead of 1 (for the dummy variables), will reflect the original scale used.

5.2 Conclusions

Based on the foregoing, the study thus, concludes that the duration of building constructions in Nigeria can be predicted through a combination of project scope factors and several other qualitative/managerial factors which influence construction time in Nigeria, namely; Cost, GFA (m2), delay in progress payment by owner, lateness in revising and approving design document by owner, delay in delivering the site to the contractor by the owner, change order by owner during construction, complexity of project design, poor site management and supervision by contractors, as well as rain effect on construction activities. However, care must be taken to ensure that proper classifications are made when developing models that can incorporate both factors to predict construction time.

5.3 Recommendations

i. This study considered building construction projects, but found that more specific classifications (public and private sector projects) produced models with better predictive abilities. It thus, recommends further studies to be conducted using other specific classification of building construction projects e.g. residential construction projects, educational projects, etc.

REFERENCES

Ameyaw, C., Mensah, S. and Arthur, Y.D. (2012). Applicability of Bromilow's time-cost model on building projects in Ghana. In: Laryea, S., Agyepong, S.A., Leiringer, R. and Hughes, W. (Eds) *Proceedings of the 4th West Africa Built Environment Research (WABER) Conference*, 24-26 July 2012, Abuja, Nigeria, 881-888.

Bromilow, F.J. (1969). Contract time performance-expectations and the reality. *Building Forum*, 1(3), 70-80.

Bromilow, F.J., Hinds, M.F. and Moody, N.F. (1980). AIQS survey of building contract time performance. *Building Economist*, 19(2), 79-82.

Chan, A.P.C. (1999). Modelling building durations in Hong Kong. *Construction Management and Economics*, 17(2), 189–96.

Chan, A.P.C. and Chan, D.W.M. (2002). Benchmarking project construction time performance: the case of Hong Kong. *Project Management-Impresario of the construction industry symposium.*

Chan, D.W.M. and Kumaraswamy, M.M. (1995). A study of the factors affecting construction duration in Hong Kong. *Construction Management and Economics*, 13(4), 319–33.

Chan, D.W.M. and Kumaraswamy, M.M. (1996). An evaluation of construction time performance in the building industry. *Building and Environment*, 31(6), 569–78.

Chan, D.W.M. and Kumaraswamy, M.M. (1999a). Forecasting construction duration for public housing projects: a Hong Kong perspective. *Building and Environment*, 34(5), 633–46.

Chan, D.W.M. and Kumarawamy, M.M. (1999b). Modelling and predicting construction duration in Hong Kong public housing. *Construction Management and Economics*, 17(3), 351–62.

Chen, W. and Huang, Y. (2006). Approximately predicting the cost and duration of school reconstruction projects in Taiwan. *Construction Management and Economics*, 24, 1231-1239.

Choudhury, I. and Phatak, O. (2004). Correlates of time overrun in commercial construction. *ASC proceedings of the 40th annual conference, Brigham Young University-Provo, Utah*.

Dascalu, C.G. and Cozma, C.D. (2008). The principal component analysis-method to reduce collinearity in multiple linear regression model; application in medical studies. *Proceedings of the 2nd WSEAS international conference on multivariate analysis and its application in Science and Engineering*.

Doguwa, S.I., and Olowofeso, O.E. (2012). Statistical analysis and reports. *In; Central Bank of Nigeria quarterly statistical bulletin*, 1(3). CBN, Garki, Abuja.

Eboh, E.C. (2009). *Social and Economic Research: Principles and Methods*. African Institute for applied Economics, Enugu, Nigeria.

Evans, R.C. and Kaka, A.P. (1998). Analysis of the accuracy of standard/average value curves using food retail building projects as case studies. *Engineering Construction and Architectural Management*, pp. 58-67.

Fellows, R. and Liu, A. (1999). *Research methods for construction*. Blackwell Science Ltd.

Field, A.P. (2005). *Discovering statistics using SPSS* (2nd edition). London; Sage.

Forsythe, P., Davidson, W. and Phua, F. (2007). Predictors of Construction time in Detached Housing projects.

Hoffman, G.J., Alfred, E., Jr, T., Webb, T.S. and Weir, J.D. (2007). Estimating performance time for construction projects. *Journal of Construction Engineering and Management, ASCE*, 193-199.

Ireland, V. (1983). The role of managerial actions in the cost, time and quality performance of high rise commercial building projects. Unpublished PhD thesis, University of Sydney, Australia.

Kaka, A., and Price, A.D.F. (1991). Relationship between value and duration of construction projects. *Construction Management and Economics*, 9(4), 384–400.

Kaka, A.P. and Price, A.D.F. (1993). Modelling standard cost commitment curves for contractors' cash flow. *Construction Management and Economics*, 11, 271-283.

Koushki, P.A. Al-Rashid, K. and Kartam, N. (2005). Delays and cost increases in the construction of private residential projects in Kuwait. *Construction Management and Economics*, 23, 285-294.

Krejcie, R. V. and Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological measurement*.

Landau, S., and Everitt, B.S. (2004). *A handbook of statistical analyses using SPSS*. Chapman and Hall/CRC press company, Washington D.C.

Leedy, P. D. (1997). *Practical research: Planning and design, 6th Edition*. Merrill, New Jersey, USA.

Long, L. and Young, D. (2009). Time-cost model of building construction projects in Korea. *Journal of Construction Engineering and Management (ICEM)*, Vol. 27, pp. 549-559.

Love, P. E. D., Tse, R. Y. C. and Edwards, D. J. (2005). Time-cost relationships in Australian building construction projects. *Journal of Construction Engineering and Management*, ASCE, 131(2), 1–8.

Maitra, S. and Yan, J. (2008). Principal Component Analysis and Partial Least Squares; two dimension reduction technique for regression. *Casualty actuarial society discussion program*.

McClave, J. T., Sincich, T., Paul, N., Carlson, W. L. and Thorne, B. (2000). *Statistics for Business and Economics* (8th Edition). New Jersey: Prentice Hall, 2000.

Mendenhall, W., Reinmuth, J.E. and Beaver, R. (1993). *Statistics for management and economics*, 7th Edition, Duxbury Press, Belmont, California, USA.

Nau, B. (2005). *Statistical forecasting methods*. Decision 411 newsgroup.

Ng, S.T., Mak, M.M.Y., Skitmore, R.M., Lam, K.C. and Varnam, M. (2001). The predictive ability of Bromilow's time-cost model. *Construction Management and Economics*, 19(2), 165–73.

Nkado, R. N. (1995). Construction time-influencing factors: the contractor's perspective. Construction *Management and Economics*, 13, 81-89.

Ogunsemi, D.R., and Jagboro, G.O. (2006). Time-cost model for building projects in Nigeria. *Construction Management and Economics*, 253-258.

Ojo, G.K. (2001). A study of the relationship between contract period and tender sum of construction projects in south-western Nigeria, unpublished MSc thesis, Obafemi Awolowo University, Ile-Ife.

Olupolola, F.O., Emmanuel, O.O., Adeniyi, M. B., and Kamaldeen, A.A. (2010). Factors affecting the time performance of building projects. *Construction Building and Real Estate Research Conference of the Royal Institution of Chartered Surveyors*, U.K

Que, B. C. (2002). Incorporating practicability into Genetic Algorithm-Based Time-Cost Optimization. *Journal of Construction Engineering and Management (ASCE)*, pp. 139-143.

Russell, B.H. (2006). *Research methods in Anthropology; Qualitative and Quantitative methods*. Walnut Creek; Altamira press.

Salkind, N. J. (1997). *Exploring research* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

Siniscalco, M.T. and Auriat, N. (1998). *Quantitative research methods in educational planning.* UNESCO international institute for educational planning.

Skitmore, R.M. and Ng, S.T. (2003). Forecast models for actual construction time and cost. *Building and Environment*, 8(8), 1075–83.

Trost, S.M. and Oberlender, G.D. (2003), Predicting accuracy of early cost estimates using factor analysis and multivariate regression, *Journal of Construction Engineering and Management*, ASCE, 129(2) 198 – 204.

Ul-Saufie, A.Z., Yahya, A.S. and Ramli, N.A. (2011). Improving multiple linear regression model using principal component analysis for predicting PM_{10} concentration in Seberancy Prai, Pulau Pinang. *International Journal of Environmental Sciences*, pp. 403-410.

Walker, D.H.T., (1995). An investigation into construction time performance. *Construction Management and Economics,* 13, 263-274.

Wuensch, K.L. (2012). Principal component analysis-SPSS. Retrieved February 15, 2013, from http://core.ecu.edu/psyc/wuenschk/wuensch.html.

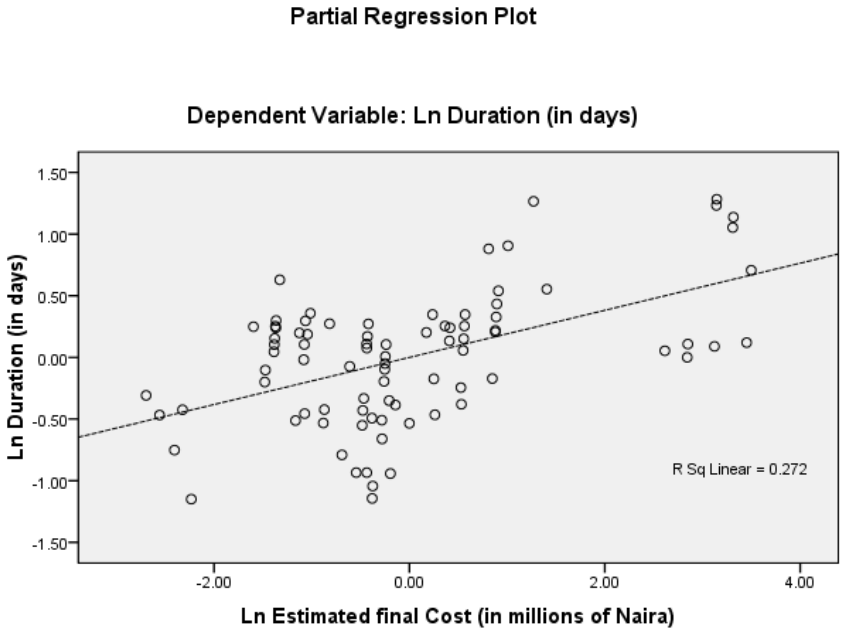
Yang, J. (2012). *Interpreting coefficients in regressions with log-transformed variables*. Cornell University, Cornell Statistical Consulting Unit.

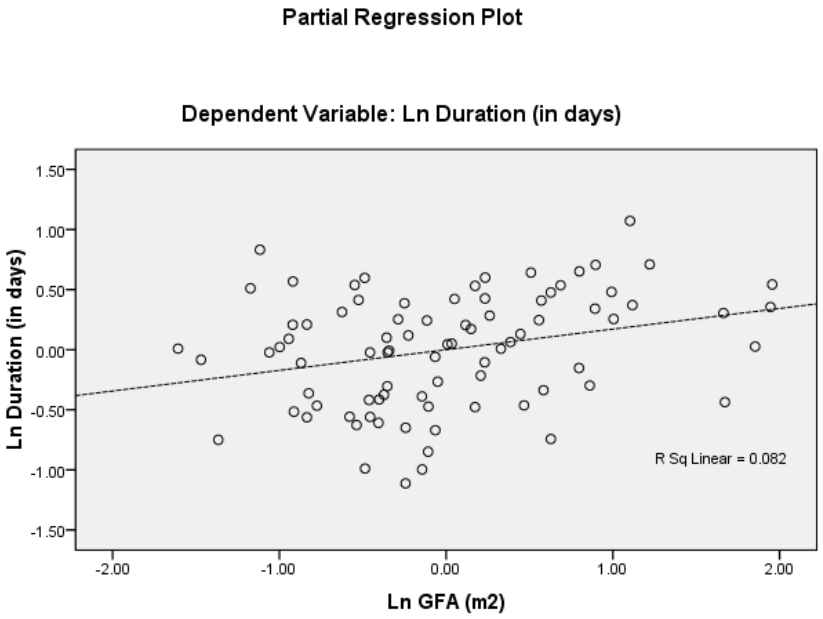
Yeong, C.M. (1994). Time and cost performance of building contracts in Australia and Malaysia, unpublished MSc thesis, University of South Australia.

APPENDIX A

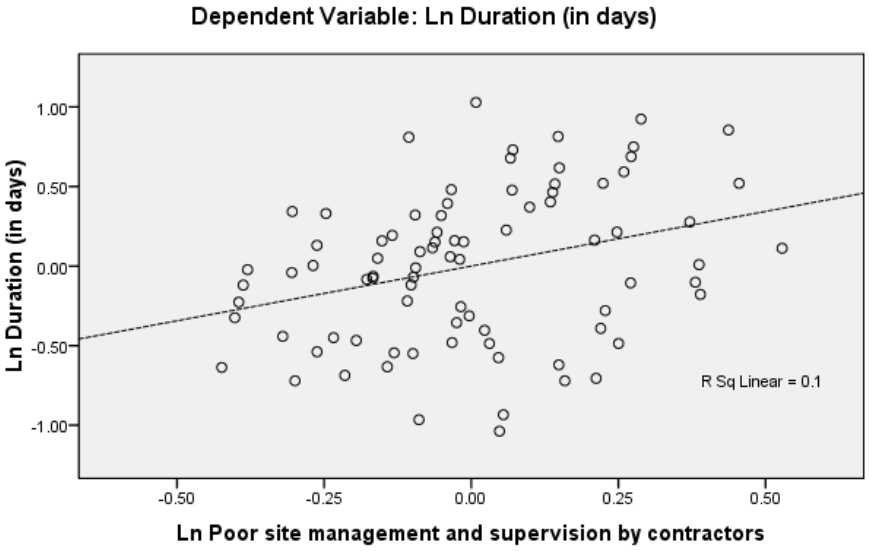
PARTIAL PLOTS

Testing for linearity (all projects):

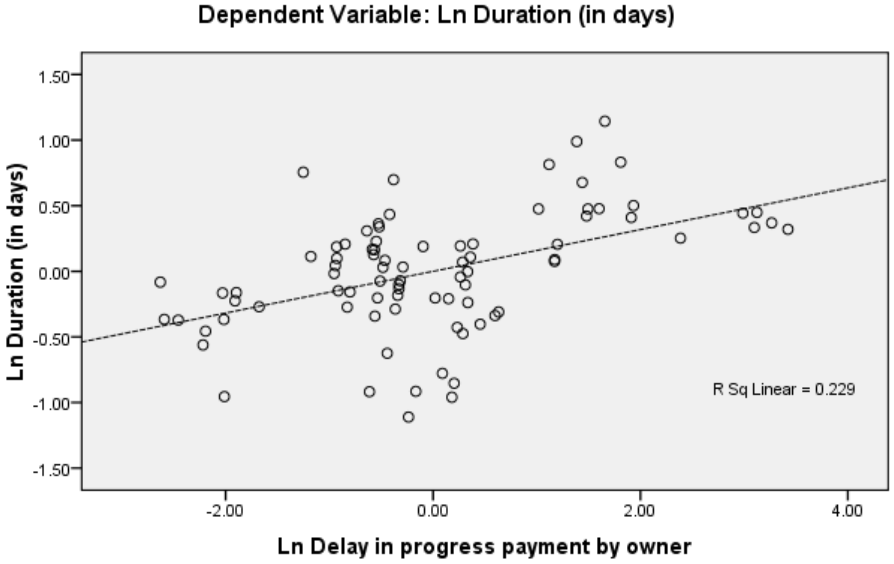




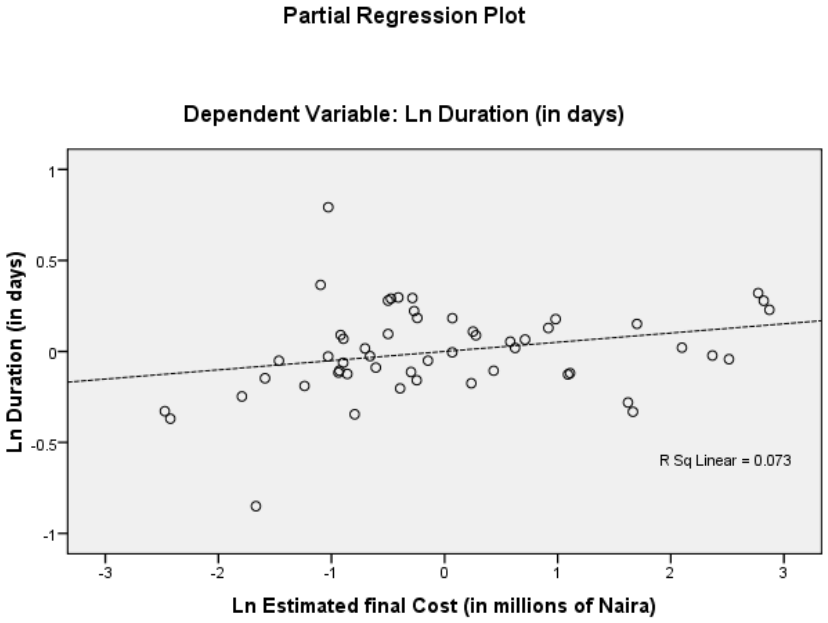
Partial Regression Plot

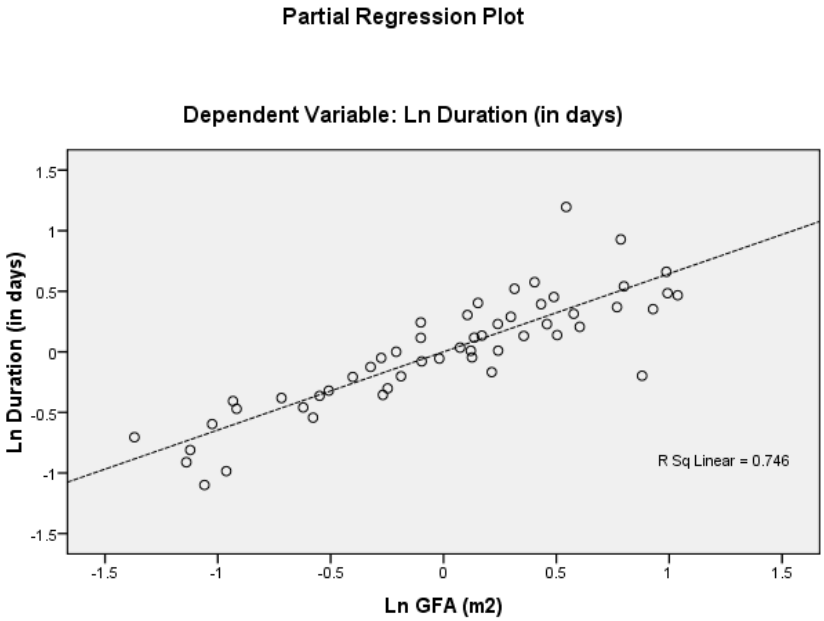


Partial Regression Plot

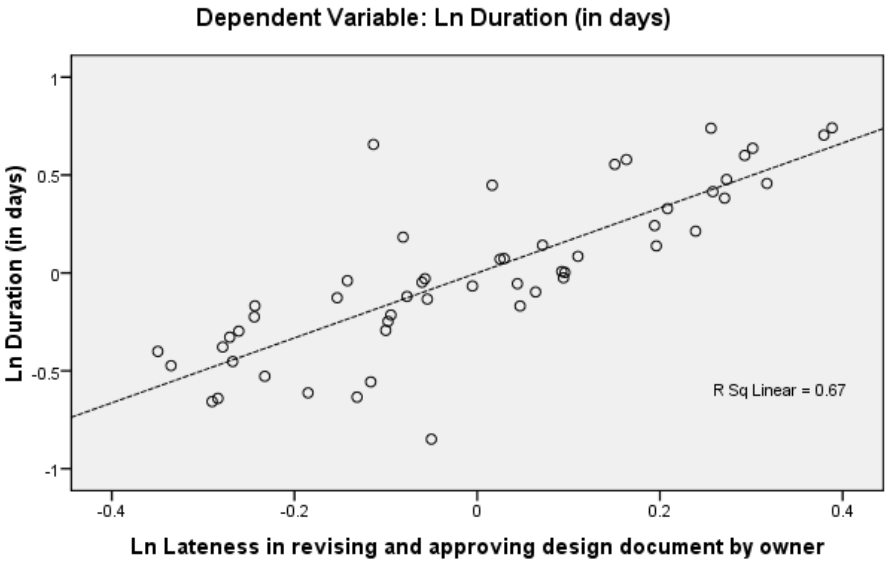


Testing for linearity (public sector projects):

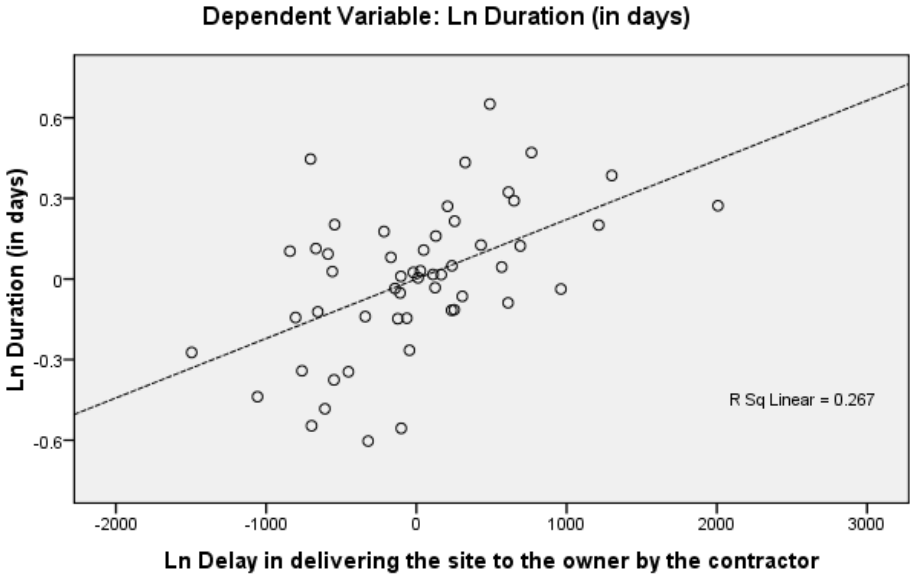




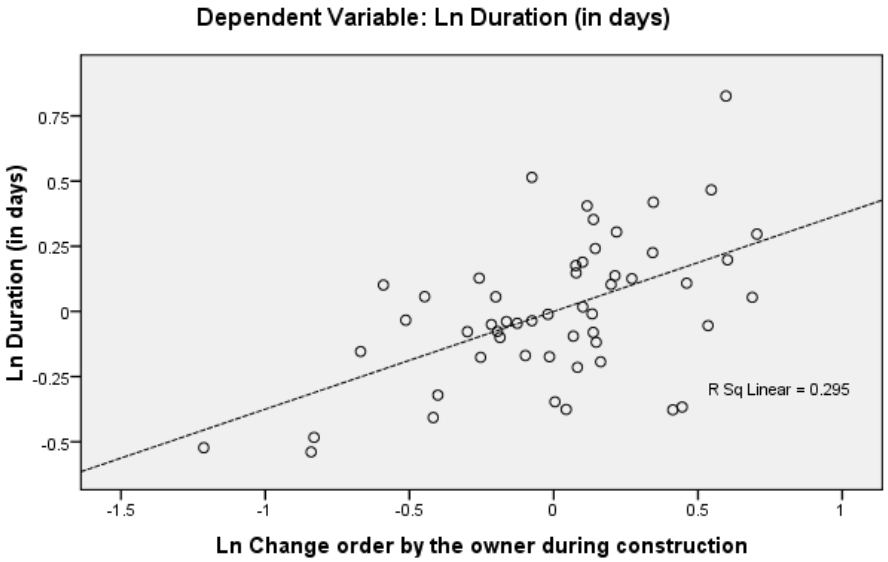
Partial Regression Plot



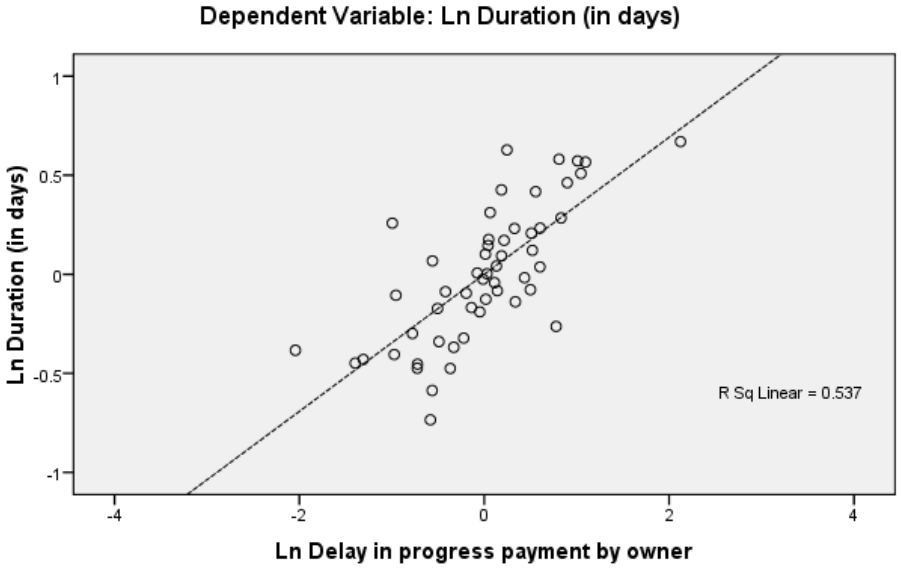
Partial Regression Plot



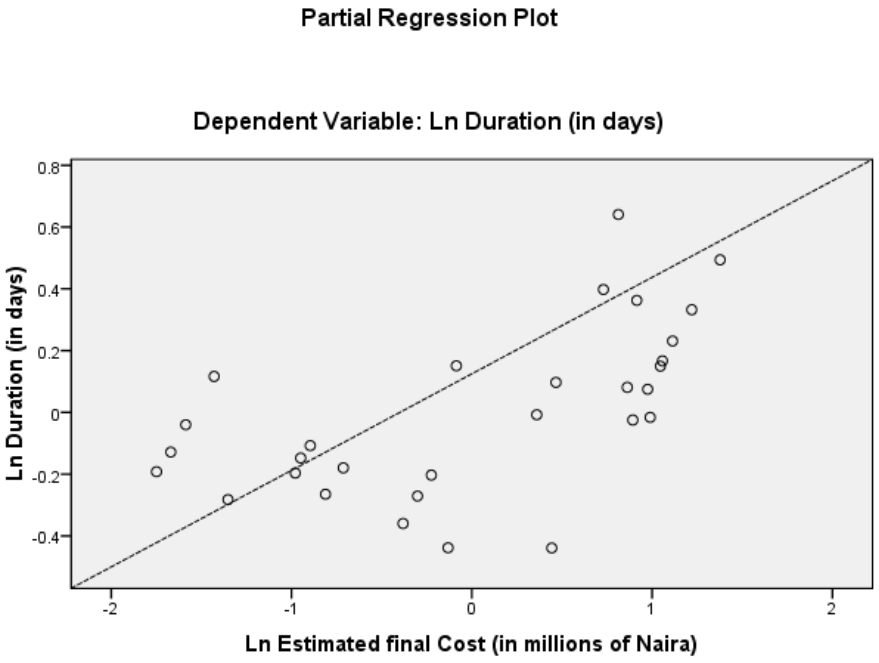
Partial Regression Plot

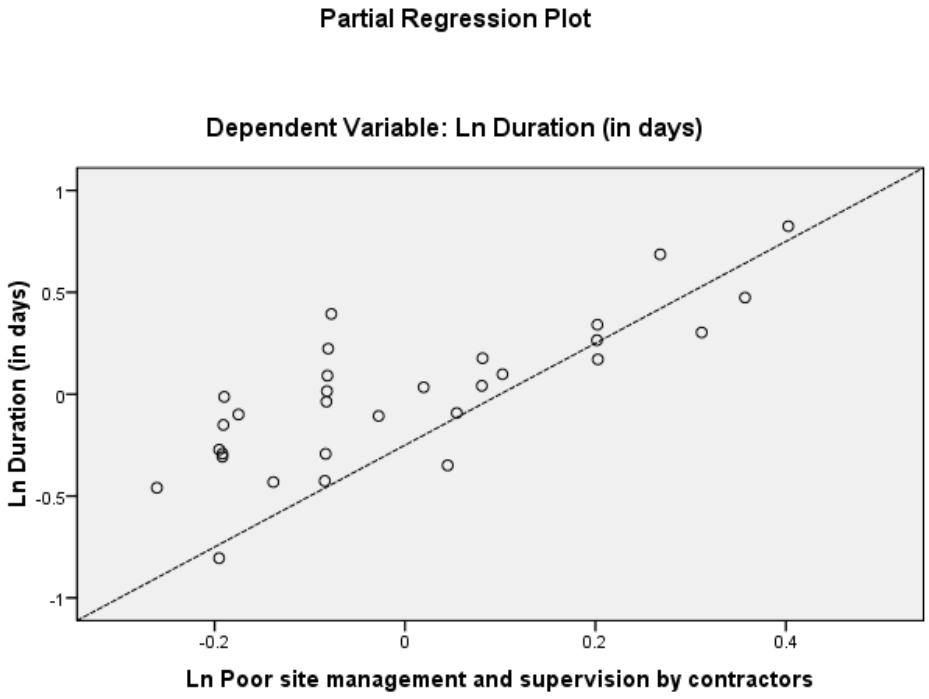


Partial Regression Plot

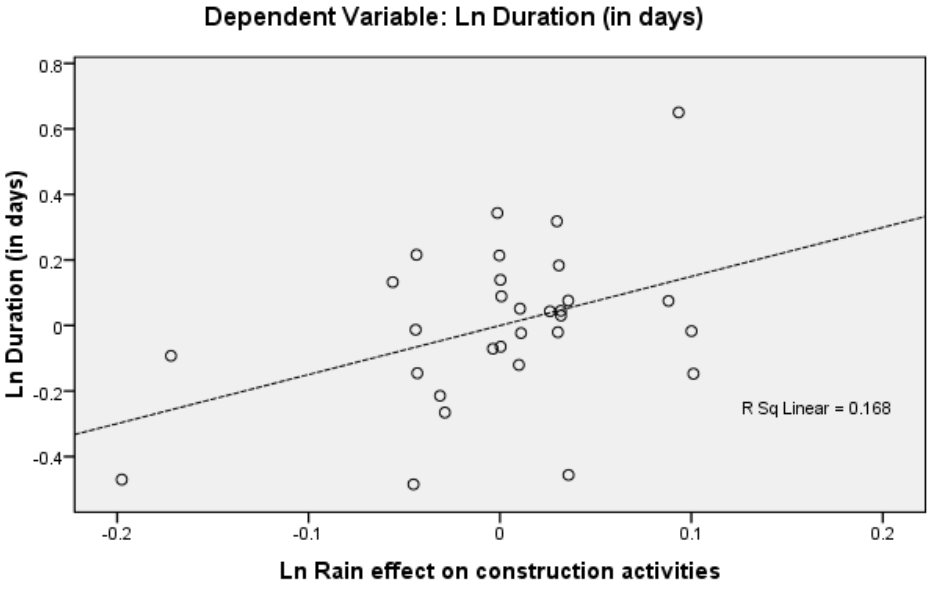


Testing for linearity (private sector projects):

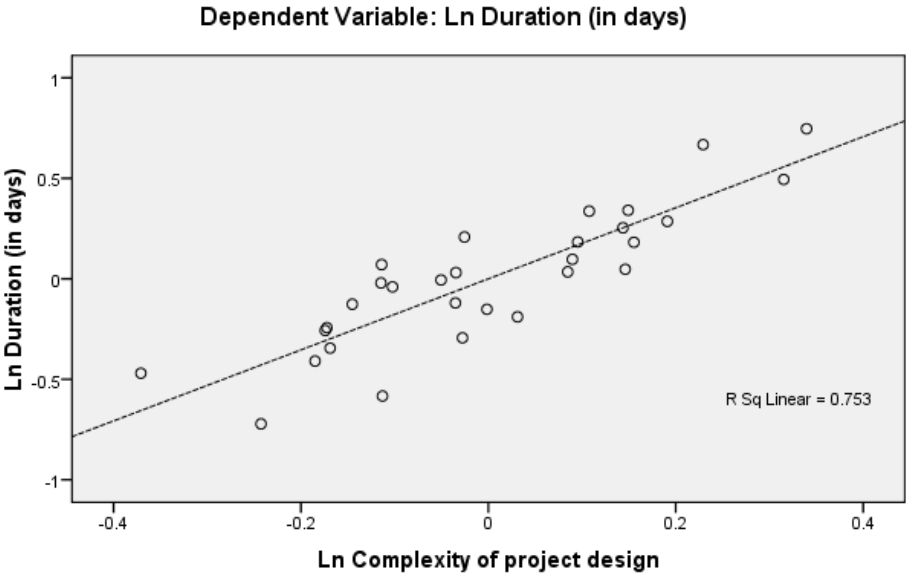




Partial Regression Plot



Partial Regression Plot



APPENDIX B

SAMPLE OF QUESTIONNAIRE